

# Die strukturelle und funktionelle Evolution des G-Protein-gekoppelten Rezeptors GPR34

Dissertation

zur Erlangung des akademischen Grades

doctor rerum medicinae

Dr. rer. med.

an der Medizinischen Fakultät

der Universität Leipzig

eingereicht von Diplom-Ernährungswissenschaftlerin Eva Engemaier

geboren am 26.10.1975 in Zittau

angefertigt am Institut für Molekulare Biochemie der Medizinischen Fakultät Leipzig

betreut von Prof. Dr. med. T. Schöneberg

Beschluss über die Verleihung des Doktorgrades vom: 24.04.2012

<b>Bibliographische Beschreibung.....</b>	<b>4</b>
<b>1 Einleitung .....</b>	<b>5</b>
1.1 <i>Das GPCR-Repertoire in Vertebraten .....</i>	5
1.1.1 Signaltransduktion.....	5
1.1.2 GPCR - Klassifikation.....	6
1.1.3 „Orphan“ G-Protein-gekoppelte Rezeptoren .....	7
1.2 <i>Die Familie der Purin- und Pyrimidinrezeptoren .....</i>	8
1.2.1 Strukturelle Besonderheiten der P2Y-Rezeptoren.....	8
1.2.2 Rolle der Nukleotidrezeptoren im kardiovaskulären System.....	9
1.3 <i>GPR34 - ein „orphan“ G-Protein-gekoppelter Rezeptor .....</i>	10
1.3.1 Gewebeexpression des GPR34.....	10
1.3.2 Genomische Organisation und Proteinstruktur des GPR34 .....	11
1.3.3 Lysophosphatidyl-L-Serin – möglicher Agonist des GPR34 .....	12
1.4 <i>GPR34-like ein verwandter „orphan“ G-Protein-gekoppelter Rezeptor.....</i>	15
<b>2 Zielstellung.....</b>	<b>16</b>
<b>3 Originalarbeit .....</b>	<b>17</b>
<b>4 Ergänzungen zur vorgenannten Originalarbeit .....</b>	<b>29</b>
4.1 <i>Vergleich der Expression von GPR34 und GPR34-like Rezeptoren im Huhn.....</i>	29
4.2 <i>Populationsgenetische Untersuchungen des humanen GPR34 .....</i>	30
<b>5 Zusammenfassung der Arbeit .....</b>	<b>31</b>
5.1 <i>Spezifische Promotoraktivität in verschiedenen Organen der Maus .....</i>	31
5.2 <i>Alternative Translationsstarts im hGPR34 .....</i>	32
5.3 <i>Geringe allelische Variabilität des GPR34-Gens in der menschlichen Population.</i>	33
5.4 <i>Ubiquitäre Expression von GPR34 und GPR34-like Rezeptoren.....</i>	34
5.5 <i>Etabliertes Mausmodell ist vital und fortpflanzungsfähig .....</i>	34

<b>6</b>	<b>Referenzen.....</b>	<b>36</b>
	<b>Anlagen.....</b>	<b>38</b>
	<b>Erklärung über die eigenständige Abfassung der Arbeit .....</b>	<b>38</b>
	<b>Eigene Publikationen .....</b>	<b>39</b>
	<b>Lebenslauf .....</b>	<b>40</b>
	<b>Danksagung.....</b>	<b>42</b>

# Bibliographische Beschreibung

Eva Engemaier

Die strukturelle und funktionelle Evolution des G-Protein-gekoppelten Rezeptors GPR34

Universität Leipzig, Dissertation

42 Seiten, 27 Literaturreferenzen, 5 Abbildungen, 4 Anlagen

Gegenstand der Arbeit ist eine umfassende Charakterisierung von Struktur und Funktion des G-Protein-gekoppelten Rezeptors GPR34 auf genomischer, mRNA und Proteinebene, die mögliche Rückschlüsse auf physiologische Funktionen oder den natürlichen Agonisten zulassen.

Dazu wurde die genomische Organisation des GPR34 in Mensch, Maus und Ratte analysiert und festgestellt, dass neben der intronlosen kodierenden Sequenz auch der 5'-Bereich des GPR34 mit seiner nicht-kodierenden Intron-Exon-Struktur stark konserviert ist. Es wurden in der Ratte und der Maus mindestens zwei, beim Menschen ein putativer Transkriptionsstart identifiziert.

Beim Menschen konnte ein kryptisches Intron innerhalb der kodierenden Sequenz des GPR34 gefunden werden, was zu einer Verkürzung des N-Terminus um 47 Aminosäuren führt.

Auf der Transkriptionsebene wurde der GPR34 und der GPR34-like Rezeptor im Haushuhn (*Gallus gallus*) und die GPR34-Expression in der Ratte mittels quantitativer RT-PCR analysiert und die ubiquitäre Gewebeverteilung des Rezeptors bestätigt.

Beim menschlichen GPR34 konnte festgestellt werden, dass die fünf putativen Translationsstarts innerhalb der kodierenden Sequenz auch als solche funktionstüchtig sind und der zweite Translationsstart bevorzugt genutzt wird. Die genetische Variabilität des GPR34 in der menschlichen Population ist sehr gering. Es konnte innerhalb einer weltweiten DNA-Probensammlung nur ein einziges Mal eine Mutation in der kodierenden Sequenz lokalisiert werden.

Mithilfe des während dieser Arbeit entstandenen Mausmodells ist eine weitere Charakterisierung der physiologischen Relevanz des GPR34 möglich.

# 1 Einleitung

G-Protein-gekoppelte Rezeptoren (GPCR) stellen die größte und vielfältigste Gruppe membranärer Rezeptoren dar. Sie sind in der Lage verschiedenste physikalische und chemische Signale wie Licht und diverse chemische Verbindungen zu detektieren. Circa 80 % aller Hormone und Neurotransmitter vermitteln ihre Signale über GPCR<sup>1</sup>. Diese Rezeptorgruppe stellt deshalb ein wichtiges medizinisches Angriffsziel dar und ist Gegenstand intensiver Forschungen. Einem Großteil der gefundenen GPCR konnte noch kein Agonist oder keine Funktion zugeordnet werden, deshalb werden diese auch als „orphan“ bzw. Waisen-Rezeptoren bezeichnet.

## 1.1 Das GPCR-Repertoire in Vertebraten

Das Genom des Menschen und anderer Säuger weist ungefähr 1000-1300 unterschiedliche Gene auf, die GPCR kodieren - beim Menschen sind dies ca. 3 bis 4 % aller Gene. GPCR sind evolutionär sehr alt und bereits in den Genomen von Protozoen, Hefen und Pflanzen zu finden. GPCR zeichnen sich durch eine gemeinsame Globalstruktur aus: die Aminosäurekette durchzieht ausgehend vom extrazellulären Raum (N-Terminus) siebenmal die Membran (Transmembranhelices) und bildet dabei drei intrazelluläre und drei extrazelluläre Schleifen (IS bzw. ES), der C-Terminus befindet sich im Zytoplasma. Die sieben Transmembranhelices (TM) sind durch einen hohen Anteil hydrophober Aminosäuren gekennzeichnet, besitzen eine  $\alpha$ -helicale Struktur und sind der am stärksten konservierte Rezeptorteil. Der extrazelluläre N-Terminus einiger GPCR ist mitunter in die Ligandenbindung involviert (z.B. bei der Familie der Glutamatrezeptoren), dieser Teil und die weiteren extrazellulären Domänen werden häufig glycosyliert. GPCR transduzieren extrazelluläre Signale vorwiegend über heterotrimere GTP-bindenden Proteine (G-Proteine). Die Übergänge von den TM in IS3, aber auch IS2 sind überwiegend an der Wechselwirkung mit G-Proteinen beteiligt.

### 1.1.1 Signaltransduktion

GPCR transduzieren extrazelluläre Signale hauptsächlich über die Aktivierung heterotrimerer G-Proteine ins Innere der Zelle. G-Proteine bestehen aus einer  $\alpha$ -Untereinheit und einer eng miteinander verbundenen dimeren  $\beta\gamma$ -Untereinheit. Bindet ein Agonist an einem GPCR, so kommt es durch Konformationsänderung des Rezeptors zur Aktivierung eines oder mehrerer G-Proteine. Das im inaktiven Zustand an die  $\alpha$ -Untereinheit des G-Proteins gebundene GDP wird gegen GTP ausgetauscht. Anschließend dissoziieren die  $G\alpha$ - und  $G\beta\gamma$ -Untereinheiten und aktivieren unterschiedliche Effektoren.

Die  $G\alpha$ -Untereinheit besitzt eine intrinsische GTPase-Aktivität, welche die Dissoziation von GTP zu GDP +  $P_i$  katalysiert. Damit schalten sich aktivierte G-Proteine wieder selbst ab. Die GDP-haltige  $G\alpha$ -Untereinheit assoziiert nun mit der  $G\beta/\gamma$ -Untereinheit und kann erneut an den Rezeptor binden - dadurch wird die wiederholte Signaltransduktion möglich.

Viele GPCR zeigen eine konstitutive Aktivität, d.h. auch ohne Bindung des Agonists findet eine Aktivierung von Signalkaskaden statt. Eine Theorie, beispielsweise vertreten von Lefkowitz et. al., besagt, dass die GPCR in einem Gleichgewicht zwischen inaktivem und aktiviertem Rezeptor stehen <sup>2</sup>. Bei einigen Rezeptoren ist das Gleichgewicht auch ohne Agonist schon so weit in Richtung des aktiven Rezeptors verschoben, dass bereits die intrazelluläre Signalkaskade aktiviert ist.

### 1.1.2 GPCR - Klassifikation

Es gibt verschiedene GPCR-Klassifikationssysteme, welche auf struktureller Ähnlichkeit der Rezeptoren, der Art der Liganden oder der Art der aktivierten G-Proteine beruhen.

Eines der GPCR-Klassifikationssysteme wurde durch Attwood und Findlay bzw. Kolakowski etabliert <sup>3,4</sup>. Es unterteilt die GPCR in sieben Familien: A bis F und O. Vriend et al. <sup>5</sup> entwickelten dieses System weiter, wobei die GPCR in nunmehr sechs Familien eingeteilt wurden: Klasse A der Rhodopsin-ähnlichen Rezeptoren, die mehr als 80 % der GPCR beinhalten, Klasse B der Secretin-ähnlichen Rezeptoren, Klasse C der metabotropen Glutamaterezeptoren, Klasse D, welche die Pheromonrezeptoren beinhaltet, Klasse E der cAMP-Rezeptoren und Klasse F der frizzled/smoothened Rezeptoren.

Eine Klassifizierung nach Zahlen von 1–6 wurde von Bockaert und Pin 1999 <sup>6</sup> vorgenommen. Die derzeit gebräuchlichste Klassifizierung der GPCR ist jedoch die nach dem GRAFS System <sup>7,8</sup>.

Das GRAFS System, welches auf phylogenetischen Untersuchungen beruht, teilt die GPCR in fünf Hauptfamilien ein: die **G**lutamat-, **R**hodopsin-, **A**dhäsions-, **F**rizzled/Taste2- und **S**ecretin-Rezeptoren. Dabei stellt die Rhodopsin-Familie (Familie A, Familie 1) die größte und vielfältigste Familie der GPCR dar und umfasst beim Menschen sowohl olfaktorische Rezeptoren (388 Gene + 479 Pseudogene), als auch nichtolfaktorische Rezeptoren (284 Gene + 27 Pseudogene) <sup>9</sup>. In diese Gruppe fallen auch die dem GPR34 ähnlichen Nukleotidrezeptoren.

Die Sequenzhomologie innerhalb der Familie A ist nicht groß, aber es existieren hochkonservierte Motive: wie z.B. das D(E)-R-Y(F)-Motiv am Übergang der TM 3 zur IS 2

und das N(D)PxxY-Motiv am Ende der TM 7. Weiterhin weisen viele Rezeptoren dieser Familie am C-Terminus ein oder zwei Cysteinreste auf, die palmitoyliert werden können und damit den C-Terminus in der Membran verankern.

### 1.1.3 „Orphan“ G-Protein-gekoppelte Rezeptoren

„Orphan“ (= Waisen) GPCR, sind Rezeptoren, deren Agonist noch nicht identifiziert werden konnte bzw. deren biologische Funktion noch nicht geklärt ist. Im Menschen sind ca. 150 der 367 identifizierten nichtolfaktorischen GPCR noch als „orphan“ Rezeptoren registriert <sup>10</sup>.

Aufgrund der großen Vielfalt der GPCR-Liganden, kann die Suche nach Agonisten für einen bestimmten Rezeptor schwierig sein. In der Vergangenheit suchte man zu einer biologisch aktiven Substanz nach dem korrespondierenden Rezeptor. Nach der Sequenzierung des humanen Genoms und verschiedener anderer Säugergenome kennt man nun das GPCR-Repertoire und sucht gezielt für „orphan“ GPCR nach Agonisten. Diese Herangehensweise wird auch als „reverse pharmacology approach“ bezeichnet. Weiterhin kann man mit Hilfe von Knock-out Strategien (z.B. Generierung eines GPCR-Gen-defizienten Mausmodells) versuchen, Informationen über die Funktion eines Rezeptors zu gewinnen. Auch Sequenzhomologien können bei der Suche nach Agonisten hilfreich sein. Dabei wird versucht von strukturell nah verwandten Rezeptoren, deren Agonist bereits bekannt ist, Rückschlüsse auf den möglichen Agonist des „orphan“ Rezeptors zu ziehen.

Einfacher als die Klassifizierung nach Liganden ist die phylogenetische Klassifizierung. So werden beispielsweise einige dieser gegenwärtig als „orphan“ bezeichneten Rezeptoren wie GPR82, GPR87, GPR171 und GPR34 durch ihre Sequenzhomologien den P2Y-Nukleotidrezeptoren zugeordnet <sup>11</sup>, obwohl deren Agonisten bisher nicht bekannt sind.

## 1.2 Die Familie der Purin- und Pyrimidinrezeptoren

Die Familie der Purin- und Pyrimidinrezeptoren sind Rezeptoren, die durch extrazelluläre Nukleoside und Nukleotide aktiviert werden. Pharmakologisch unterscheidet man P1 (Adenosin) und P2 (Adenosintriphosphat - ATP und Adenosindiphosphat - ADP) Rezeptoren<sup>12</sup>. Die P2-Rezeptoren werden in zwei weitere Gruppen unterteilt: die ionotropen P2X und die metabotropen P2Y-Rezeptoren. Ionotrope P2X-Rezeptoren sind durch Nukleotide aktivierbare Ionenkanäle. Die P1- und metabotropen P2Y-Rezeptoren gehören zur Gruppe der rhodopsinähnlichen GPCR und letztere sind strukturell dem GPR34 am ähnlichsten. Agonisten können sowohl Purin- als auch Pyrimidinderivate sein<sup>12</sup>.

### 1.2.1 Strukturelle Besonderheiten der P2Y-Rezeptoren

Bisher konnten acht verschiedene Säuger-P2Y-Rezeptorsubtypen identifiziert werden (P2Y<sub>1, 2, 4, 6, 11, 12, 13, 14</sub>), die, außer dem P2Y<sub>11</sub>-Rezeptor, durch intronlose Gene kodiert werden. Anhand ihres affinsten Agonisten werden sie in vier Untergruppen eingeteilt: Adenin-Nukleotid-aktivierte Rezeptoren (P2Y<sub>1</sub>, P2Y<sub>11</sub>, P2Y<sub>12</sub>, P2Y<sub>13</sub>), Uracil-Nukleotid-aktivierte Rezeptoren (humaner P2Y<sub>4</sub>, P2Y<sub>6</sub>), durch Purin- und Pyrimidinnukleotide aktivierte Rezeptoren (P2Y<sub>2</sub>, P2Y<sub>4</sub>) und UDP-Glucose bzw. UDP-Galaktose aktivierte Rezeptoren (P2Y<sub>14</sub>)<sup>12</sup>. Allerdings ist diese Einteilung nach Agonisten schwierig, da viele der Rezeptoren durch mehrere Nukleotide aktivierbar sind und neben den Nukleotiden auch andere Liganden bei strukturell ähnlichen Rezeptoren identifiziert werden konnten, so dass Nukleotide möglicherweise nicht die einzigen Agonisten für diese Rezeptorgruppe darstellen<sup>13-15</sup>.

Deshalb wurde von Schöneberg et al. 2007<sup>11</sup> vorgeschlagen, diese Rezeptorgruppe eher nach phylogenetischen Gesichtspunkten als nach Ligandenspezifität zu klassifizieren. Anhand ihrer Aminosäuresequenzähnlichkeit werden die P2Y-Rezeptoren in zwei Gruppen unterteilt: die P2Y<sub>1, 2, 4, 6, 11</sub> und die P2Y<sub>12, 13, 14</sub>-Gruppe<sup>11</sup>. Die humanen P2Y-Rezeptoren sind innerhalb der TM1-7 zu 21 bis 48 % konserviert<sup>12</sup>. Die höchste Aminosäure-Sequenzidentität findet man innerhalb der P2Y<sub>12, 13, 14</sub>-Gruppe (47 bis 48 %).

Der N-Terminus der metabotropen Nukleotidrezeptoren ist relativ kurz und scheint keine spezifische Funktion zu haben. Alle humanen P2Y-Rezeptoren besitzen das HXXR/K-Motiv in der TM6 (siehe Abb. 1), was für die Bindung der Agonisten wichtig sein könnte. Das Q/KXXR Motiv in ES3 ist möglicherweise ebenfalls an der Ligandenbindung innerhalb der P2Y<sub>1, 2, 4, 6, 11</sub>-Gruppe beteiligt. Es ist in der P2Y<sub>12, 13, 14</sub>-Gruppe durch das KEXXL-Motiv ersetzt, was Auswirkungen auf die Art der Ligandenbindung haben könnte<sup>12</sup>.



Die Aktivierung der P2Y-Rezeptoren durch Nukleotide führt zur Aktivierung der Phospholipase C und Freisetzung von intrazellulärem Calcium bzw. zur Beeinflussung der Adenylylzyklase und somit zu einem veränderten cAMP-Spiegel in der Zelle.

Die Rezeptoren P2Y<sub>1, 2, 4, 6, 11</sub> koppeln hauptsächlich an G $\alpha_{q/11}$ -Proteine, P2Y<sub>12, 13, 14</sub> sind mit G $\alpha_{i/o}$ -Proteinen assoziiert. Einzelne Rezeptoren können aber zusätzlich auch andere Signalwege aktivieren, so z.B. den MAP (mitogen-aktivierte Proteinkinase)-Kinase-Weg<sup>12,16</sup>.

### 1.2.2 Rolle der Nukleotidrezeptoren im kardiovaskulären System

Nukleotidrezeptoren sind in vielen verschiedenen Zellen exprimiert (Nervenzellen, Muskelzellen, Zellen des Immunsystems). Den beiden Rezeptoren P2Y<sub>1</sub> und P2Y<sub>12</sub> wird im kardiovaskulären System eine besondere Position eingeräumt. Sie spielen eine bedeutende Rolle bei der ADP-induzierten Thrombozytenaggregation. Dabei initiieren sie die Aktivierung, die Änderung der Form und die nachfolgende kurzzeitige Aggregation der Blutplättchen durch ADP oder Kollagen. Zusätzlich amplifizieren und stabilisieren sie die Thrombozytenaggregation, welche auch durch andere Signalmoleküle wie TXA<sub>2</sub>, Thrombin, Kollagen, Chemokine und Immunkomplexe ausgelöst werden kann und bilden den Angriffspunkt für den Thrombozytenaggregationshemmer Clopidogrel<sup>12</sup>. Clopidogrel ist ein Prodrug. Ein Metabolit des Clopidogrels bindet kovalent an den P2Y<sub>12</sub>-Rezeptor, inaktiviert ihn irreversibel durch Spaltung der strukturstabilisierenden Disulfidbrücken und ist damit ein potenter Thrombozytenaggregationshemmer. P2Y<sub>12</sub>-defiziente Mäuse zeigen genauso wie Clopidogrel-behandelte Mäuse eine Hemmung der Thrombozytenaggregation und damit eine verlängerte Blutungszeit.

P2Y<sub>1</sub>- bzw. P2Y<sub>2</sub>-Rezeptoren werden ebenfalls in Gefäßendothelzellen exprimiert, wo sie in die Freisetzung von NO und die nachfolgende Vasodilatation sowie in die Zellproliferation und die Expression von Adhäsionsproteinen für Monozyten involviert sind.

### 1.3 GPR34 - ein „orphan“ G-Protein-gekoppelter Rezeptor

Der GPR34 wurde gleichzeitig von zwei Arbeitsgruppen - Marchese und Schöneberg 1999 als „orphan“ GPCR unabhängig voneinander identifiziert <sup>17,18</sup>. Er gehört zur Gruppe der rhodopsinähnlichen Rezeptoren und zeigt die größte Aminosäuresequenz-Homologie zu den P2Y-Rezeptoren. Der humane GPR34 wird durch eine Kette von 381 Aminosäuren (AS) gebildet. In anderen Spezies variiert diese Länge zwischen 369 und 389 Aminosäuren. Der GPR34 konnte in allen untersuchten Vertebratenklassen identifiziert werden: in Säugern, Beuteltieren, Vögeln, Amphibien, Reptilien, Knochen- und Knorpelfischen <sup>19</sup>. Da er in Knorpelfischen aber nicht in niederen Wirbeltierklassen (Schleimaal, Neunauge) oder Insekten vorkommt (in NCBI-Genomdatenbanken für *Caenorhabditis elegans*, *Drosophila melanogaster* und *Anopheles gambiae* ist der GPR34 nicht zu finden), ist anzunehmen, dass der Rezeptor vor mindestens 450 Millionen Jahren entstanden ist. Über diesen Zeitraum wurden nur 17 % aller Aminosäuren des GPR34 zu 100% evolutionär konserviert. Im Vergleich dazu sind über denselben Zeitraum 43% der Aminosäuren des Rhodopsins konserviert <sup>19</sup>.

Der GPR34 wird an der Zelloberfläche exprimiert. Er hemmt die Forskolin-induzierte cAMP-Bildung und zeigt eine erhöhte Basalaktivität im  $G_{\alpha i}$ -Signalweg. Diese Ergebnisse konnten zusätzlich in einem Phosphoinositol-Assay unter Zuhilfenahme eines chimären G-Proteins ( $G_{\alpha \Delta 6qi4myr}^{20}$ ) bestätigt werden. Obwohl damit bekannt war, dass der GPR34 sein Signal über die Kopplung an  $G_{\alpha i}$ -Proteine transduziert, konnte mit Hilfe des „reverse pharmacology approach“ durch Testen von großen Substanzbibliotheken an GPR34-exprimierenden Zellen bis zum Beginn dieser Promotionsarbeit kein Agonist identifiziert werden <sup>18,19</sup>.

#### 1.3.1 Gewebeexpression des GPR34

Mittels „Northern Blot“ konnte die Expression des GPR34 in Niere, Milz, Herz, Hirn, Plazenta, Lunge, Leber, Skelettmuskel, Pankreas, Thymus, Testes, Eierstöcken, Dünndarm, Kolon und in Blutzellen des Menschen festgestellt werden. Das gleiche Expressionsmuster wurde in der Maus und der Ratte durch Northern Blot nachgewiesen. Diese ubiquitäre Verteilung des Rezeptors könnte auf die Expression in Zellen (Immunzellen, Gefäßzellen) zurückzuführen sein, die in nahezu allen Geweben vorkommen <sup>19</sup>.

### 1.3.2 Genomische Organisation und Proteinstruktur des GPR34

Der GPR34 ist im menschlichen Genom X-chromosomal lokalisiert (Xp11.3–11.4) und befindet sich in seiner DNA-Sequenz als einzelne Kopie in Antisense-Orientierung innerhalb eines Introns einer Calcium/Calmodulin-abhängigen Serinproteinkinase. Seine kodierende Sequenz ist in allen Säugern intronlos. Das für den GPR34 aus der Anzahl der Aminosäuren errechnete Molekulargewicht beträgt 44 kDa. Im „Western Blot“ zeigte sich eine ca. 75–90 kDa große Proteinbande, die auf eine Glycosylierung des GPR34 hindeutet<sup>18</sup>. In den Fischen (z.B. Karpfen und Zebrafisch) wurden zwei Typen des GPR34 gefunden, charakteristische Motive für die jeweiligen Subtypen konnten jedoch noch nicht identifiziert werden<sup>19</sup>.

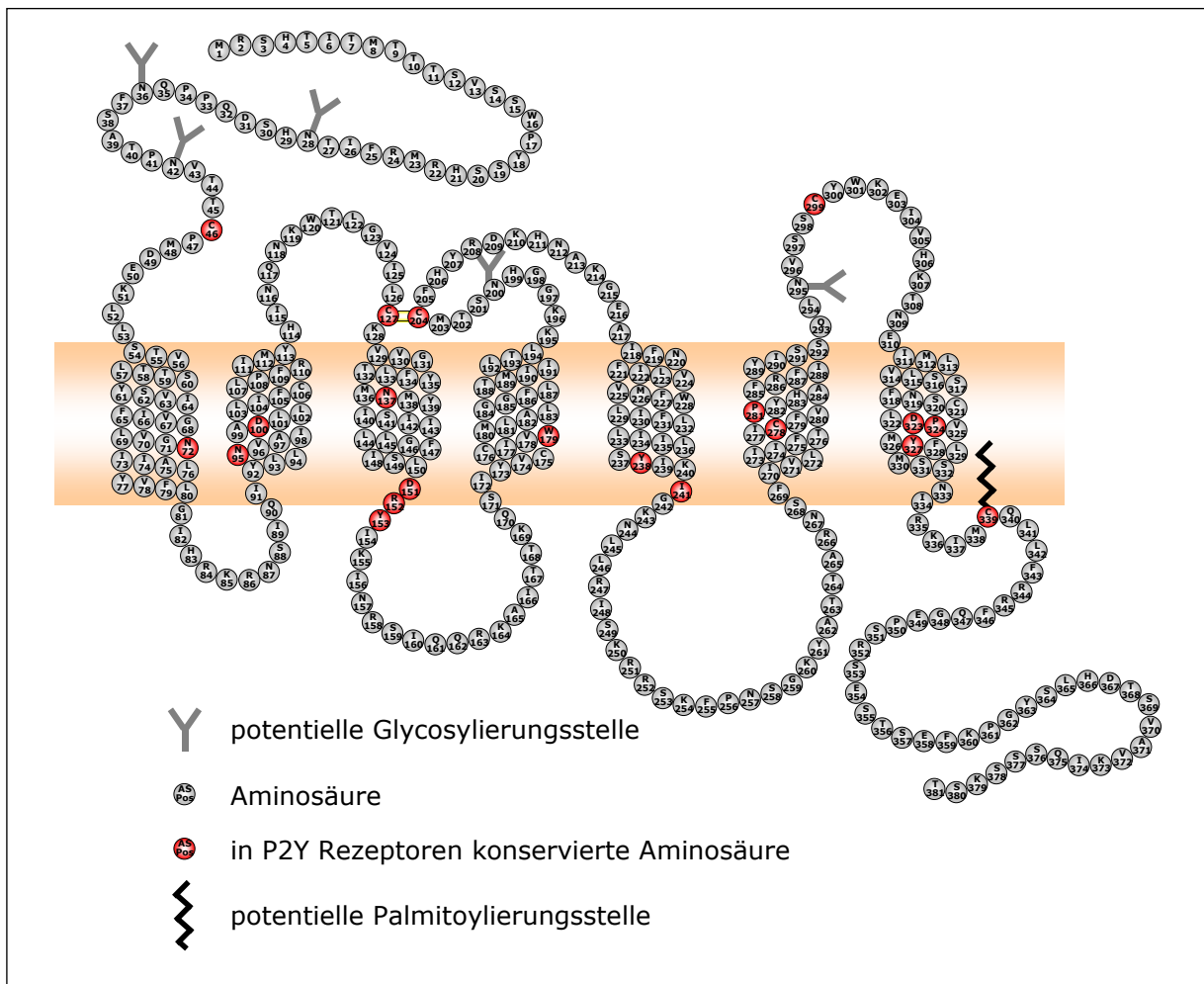


Abbildung 1: Aminosäuresequenz und Struktur des humanen GPR34. Gekennzeichnet sind die konservierten Aminosäuren, die potentiellen Glycosylierungsstellen und die potentielle Palmitoylierungsstelle.

Wie in Abbildung 1 gezeigt, finden sich im humanen GPR34 fünf potentielle N-Glycosylierungsstellen (NXS/T) im extrazellulären N-Terminus (Asn<sup>28</sup>, Asn<sup>36</sup>, Asn<sup>42</sup>), in der ES2 (Asn<sup>200</sup>) und der ES3 (Asn<sup>295</sup>), sie sind allerdings nicht in allen Spezies konserviert. Die

beiden Cysteine an Position 127 und 204 in der IS1 und der IS2 bilden wie in anderen rhodopsinähnlichen Rezeptoren eine Disulfidbrücke zur Stabilisierung der Rezeptorstruktur. Zwei zusätzlich im GPR34 enthaltene und konservierte Cysteine (Cys<sup>42</sup>, Cys<sup>299</sup>) könnten ebenfalls an der Bildung einer Disulfidbrücke beteiligt sein, die den N-Terminus mit dem ES3 verbindet. Allerdings sind diese beiden im Spiegelkärpfling und Rotbarsch nicht konserviert<sup>19</sup>. Ähnliches gilt für die Palmitoylierungsstelle nahe dem C-Terminus (Cys<sup>339</sup>). Sie ist außer im GPR34 des Platypus, Karpfen Subtyp 1, Tetraodon und Fugu in 62 anderen untersuchten Orthologen konserviert. Die in den meisten Rhodopsin-ähnlichen Rezeptoren konservierten Aminosäuren sind auch in den bisher untersuchten GPR34 Orthologen zu finden: Asn<sup>95</sup> in TM2, das DRY-Motiv am C-terminalen Ende von TM3, das Tyr<sup>238</sup> in TM5 und das Pro<sup>281</sup> in TM6<sup>18</sup>. Allerdings ist in Fugu und Tetraodon das Asp im DRY-Motiv durch ein His ersetzt. Es ist also hier möglicherweise für den GPR34 nicht funktionell relevant, oder die Funktion des GPR34 wird in diesen Tierarten durch andere Mechanismen kompensiert. Weiterhin ist in TM7 das Asn durch ein Asp beim sonst in allen Rhodopsin-ähnlichen Rezeptoren vorhandenen NPXXY-Motiv ersetzt.

Im Vergleich der TM des GPR34 sind die TM6 und TM7 stark konserviert. So ist z.B. das SXXCFXY-Motiv in der ES2 und das FPNXXXKYXXXTAXXSf-Motiv in der IS3 des GPR34 einzigartig, verglichen mit anderen Rhodopsin-ähnlichen GPCR. Das E/S-S-T-S/T-E-X-K/R-Motiv ist in den meisten C-Termini des GPR34 zu finden. Es ähnelt dem Substratmotiv für eine Caseinkinase, ist jedoch nicht in den Fischen im GPR34 Subtyp 2 enthalten<sup>19</sup>.

Obwohl der GPR34 die größte Homologie zu den P2Y-Rezeptoren zeigt, schließt Marchese die Aktivierung des Rezeptors durch Nukleotide nahezu aus, weil der GPR34 nicht die geladene basische Aminosäure innerhalb der TM7 besitzt, die essentiell für die Aktivierbarkeit durch Nukleotide sein soll<sup>17</sup>. Auch Untersuchungen der eigenen Arbeitsgruppe konnten bisher keine Nukleotide als Agonisten des GPR34 identifizieren<sup>18</sup>.

### 1.3.3 Lysophosphatidyl-L-Serin – möglicher Agonist des GPR34

Sugo et al. identifizierten 2006 Lysophosphatidyl-L-Serin (1-Palmitoyl-*sn*-glycero-3-phosphoridyl-L-serin = LysoPS) als Agonisten für den GPR34<sup>21</sup>. Dazu wurde eine den humanen GPR34 stabil exprimierende CHO (chinese hamster ovary)-Zelllinie mit verschiedenen Fraktionen von Gewebeextrakten versetzt. Bestimmte Extrakte von Ratten- und Schweinehirn zeigten dabei eine Hemmung der Forskolin-stimulierten cAMP-Bildung. Da die Hemmung auch nach Protease-Vorbehandlung noch nachweisbar war, konnte ausgeschlossen werden, dass es sich bei dem Agonisten um ein Protein handelte. LysoPS

wurde als Agonist für den GPR34 beim Screening von Substanzbibliotheken identifiziert. Es hemmte signifikant die Forskolin-stimulierte cAMP-Bildung, mit einem  $EC_{50}$ -Wert von  $270 \pm 32$  nM. Dagegen waren keine dem LysoPS ähnlichen Substanzen wie Lysophosphatidylcholin, Lysophosphatidylethanolamin oder Lysophosphatidylsäure als Agonisten am GPR34 wirksam. Die durch LysoPS ausgelöste Hemmung der Forskolin-stimulierten cAMP Bildung war Pertussistoxin-sensitiv, womit bestätigt werden konnte, dass der Rezeptor an  $G_{\alpha i}$ -Proteine koppelt. Auch in Maus- und Meerschwein-GPR34 stabil exprimierenden Zelllinien konnte LysoPS als Agonist ( $EC_{50}$ :  $49 \pm 4,5$  nM bzw.  $240 \pm 36$  nM) bestätigt werden.

Weiterhin konnten Sugo et al. in Mastzellen von Mensch, Maus und Ratte eine hohe Expression des GPR34 nachweisen<sup>21</sup>. An peritonealen Mastzellen der Ratte konnten sie nach Zugabe von 100 nM LysoPS eine stark beschleunigte, durch Antigene oder Nervenwachstumsfaktor (Nerve Growth Factor, NGF) induzierte Degranulation der Zellen beobachten. In der menschlichen Mastzelllinie LAD2 konnte auch eine Pertussistoxin-sensitive Degranulation allein durch LysoPS festgestellt werden. Diese Ergebnisse müssen allerdings durch weitere Experimente bestätigt werden.

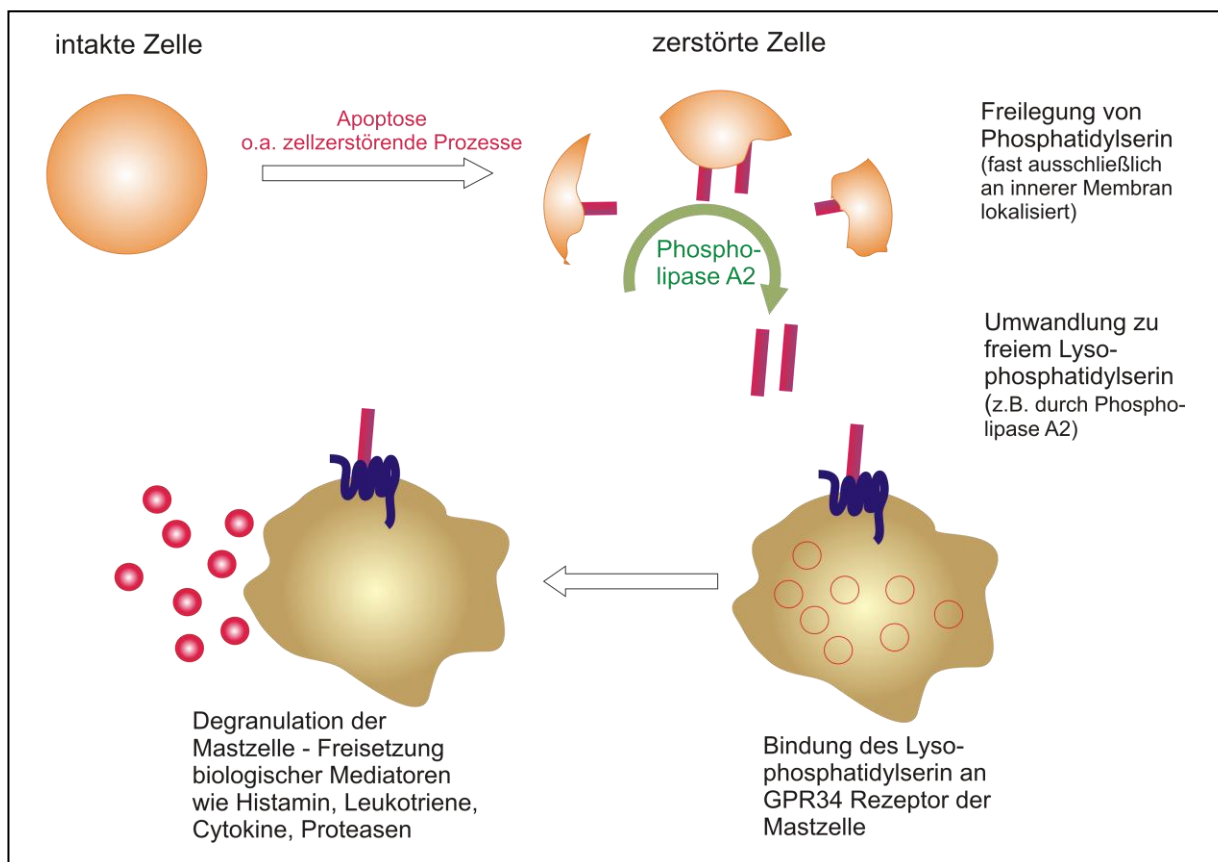


Abbildung 2: Hypothetischer Mechanismus der Degranulation von peritonealen Mastzellen der Ratte durch Aktivierung des GPR34 mittels Lysophosphatidyl-L-Serin<sup>21,22</sup>.

LysoPS wird als immunologischer Regulator betrachtet, da es die Degranulation von peritonealen Mastzellen der Ratte beschleunigt. Gelangt beispielsweise ein Antigen in die Blutbahn werden die Mastzellen aktiviert. Daraufhin degranulieren sie und setzen Signalmoleküle wie Histamin, Leukotriene, Prostaglandine, Proteasen und Cytokine frei. Dies spielt insbesondere beim Auslösen von Allergien eine Rolle. Die durch LysoPS erhöhte Histaminfreisetzung ist mit Pertussistoxin teilweise unterdrückbar. Dies deutet auf den Einfluss eines  $G_{\alpha i}$ -gekoppelten GPCR hin <sup>23</sup>.

Die Vorstufe von LysoPS, das Phosphatidylserin, ist ausschließlich an der inneren Zellmembran lokalisiert. Für die Aktivierung der Degranulation der Mastzellen muss LysoPS im Extrazellularraum vorliegen. Vermutlich erfolgt die Bildung von LysoPS aus dem Phosphatidylserin nach außen gekehrter Zellmembranen, die durch Apoptose entstehen. Deshalb wurde LysoPS bereits als mögliches Signalmolekül für apoptotische Prozesse benannt <sup>24</sup> (siehe Abbildung 2).

## 1.4 GPR34-like ein verwandter „orphan“ G-Protein-gekoppelter Rezeptor

Wie bereits in der Arbeit von Schulz 2003<sup>19</sup> gezeigt, konnte in Knochen- und Knorpelfischen, Vögeln, aber nicht in Säugern ein naher Verwandter des GPR34 nachgewiesen werden – der GPR34-like. Er ist ebenfalls ein „orphan“ Rezeptor und stellt in seiner Aminosäuresequenz ein Bindeglied zwischen GPR34 und der Gruppe der P2Y-Rezeptoren dar (Abbildung 3).

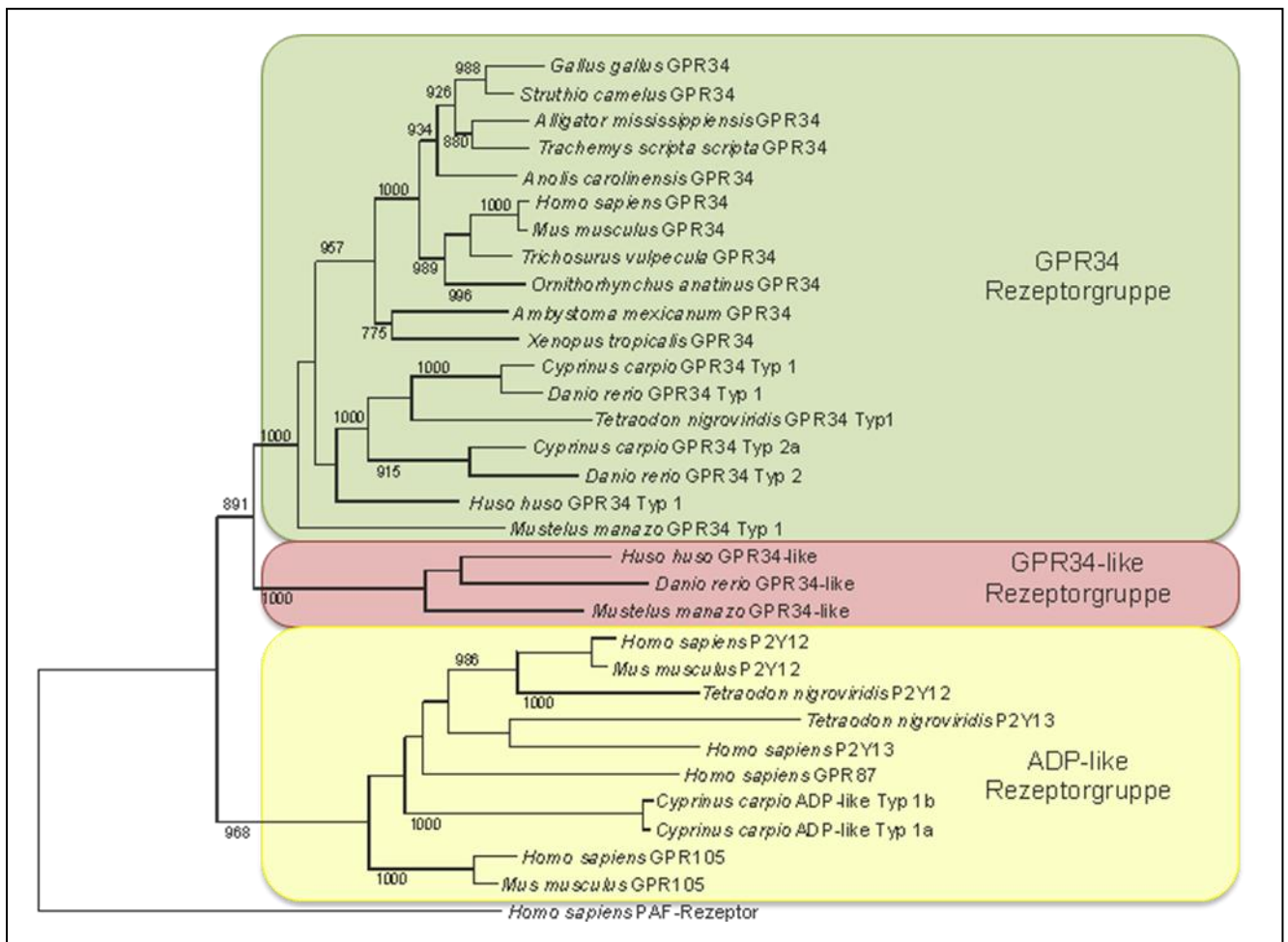


Abbildung 3: Phylogenetische Verwandtschaft zwischen GPR34-, GPR34-like- und P2Y-Rezeptoren. Die Aminosäuresequenz der genannten Rezeptoren wurde mittels Clustal X analysiert. Die Zahlen stellen die berechneten Bootstrap-Werte dar. Nur statistisch gesicherte Bootstrap-Werte über 750 wurden berücksichtigt (nach Schulz et. al. 2003<sup>19</sup>).

Phylogenetisch ist der GPR34-like der Gruppe der P2Y-Rezeptoren zuzuordnen. Leider gibt es noch keine weiteren Erkenntnisse über seinen Agonisten oder seine physiologische Relevanz.

## **2 Zielstellung**

Ziel der Arbeit war eine umfassende Charakterisierung von Struktur und Funktion des G-Protein-gekoppelten Rezeptors GPR34 auf genomischer, mRNA und Proteinebene, die mögliche Rückschlüsse auf physiologische Funktionen oder den natürlichen Agonist zulassen. Dazu sollten folgende Teilziele erreicht werden:

1. Es sollte die genomische Organisation des GPR34 in verschiedenen Vertebraten geklärt werden.
2. Es sollte untersucht werden, ob es Splice-Varianten des GPR34 in Säugern gibt. Gleichzeitig sollte der Transkriptionsstart für nachfolgende Promotoranalysen identifiziert werden.
3. Es sollte auf quantitativer Ebene das Expressionsprofil in verschiedenen Spezies verglichen werden. Im Falle von Splice- oder Transkriptvarianten sollten diese gegeneinander quantifiziert werden, um die physiologisch relevanten Transkripte zu identifizieren.
4. Die bereits in Orthologvergleichen gefundene hohe Variabilität der Länge des N-Terminus des GPR34 sollte detailliert untersucht und der Translationsstart bestimmt werden.
5. Es sollte die Variabilität des GPR34 in der menschlichen Population bestimmt und eventuelle Varianten auf ihre funktionelle Relevanz überprüft werden.
6. Es sollte ein Vergleich in den Expressionsprofilen des GPR34 und des GPR34-like im Huhn vorgenommen werden, um anhand der Expressionsprofile eine mögliche physiologische Relevanz zu erkennen.

Die Ergebnisse dieser Untersuchungen sollten den Kenntnisstand über die physiologische Relevanz des GPR34 im Menschen und in anderen Vertebraten erweitern.



### **3 Originalarbeit**

#### **Genomic and supragenomic structure of the nucleotide-like G protein-coupled receptor GPR34**

#### **Genomische und supragenomische Struktur des Nukleotidrezeptor-ähnlichen G-Protein-gekoppelten Rezeptors GPR34**

Viele neue Mitglieder der G-Protein-gekoppelten Rezeptoren (GPCR) wurden durch direkte Klonierungsversuche bzw. breit angelegte Sequenzierungen von Wirbeltiergenomen entdeckt. Zu diesen neu entdeckten GPCR gehört ebenfalls der GPR34, bei dem initiale Studien zeigen konnten, dass er evolutionär konserviert wurde und strukturell der ADP-Rezeptor P2Y<sub>12</sub>-ähnlichen GPCR-Gruppe zugeordnet werden kann. Für eine genauere Charakterisierung wurde die genomische Organisation, die Regulierung der Expression und die supragenomische Diversität des GPR34-Rezeptors in diversen Wirbeltierarten analysiert.

Im Gegensatz zur intronlosen Kodierungssequenz konnte beim GPR34 eine evolutionär konservierte 5'-Intron-Exon-Struktur identifiziert werden. Ebenfalls wurde ein alternativ genutztes kryptisches Intron innerhalb der Kodierungssequenz gefunden, welches den N-Terminus um 47 Aminosäuren verkürzt.

Die ubiquitäre Verteilung des Rezeptors wird durch genomische Sequenzen 5' von mindestens zwei Transkriptionsstartregionen in Ratten und Mäusen, bzw. einer Transkriptionsstartregion im Menschen realisiert. In Nagern werden beide Promotoren in allen untersuchten Geweben genutzt, dabei ist der Grad der Aktivität gewebespezifisch.

Mehrere konservierte, im Leserahmen befindliche AUGs innerhalb der ersten 150 bp der Kodierungssequenz können als Translationsstart im Menschen und anderen Säugern dienen. Kombinierte Mutagenese und die Expression von Reportergenkonstrukten bestätigten diese multiplen Translationsstartpunkte und zeigten eine Präferenz für das zweite im Leserahmen liegende AUG des humanen GPR34. Unsere Ergebnisse bestätigen, dass multiple Translationsstarts und alternatives Spleißen zur Rezeptorvielfalt des GPR34 beitragen.

# Genomic and supragenomic structure of the nucleotide-like G-protein-coupled receptor GPR34<sup>☆</sup>

Eva Engemaier, Holger Römpler, Torsten Schöneberg, Angela Schulz<sup>\*</sup>

*Institute of Biochemistry, Department of Molecular Biochemistry, Medical Faculty, University of Leipzig, Johannisallee 30, 04103 Leipzig, Germany*

Received 8 July 2005; accepted 7 October 2005

Available online 9 December 2005

## Abstract

Directed cloning approaches and large-scale sequencing of several vertebrate genomes unveiled many new members of the G-protein-coupled receptor (GPCR) superfamily, among them GPR34. Initial studies showed that GPR34 is an evolutionarily old GPCR structurally related to a group of ADP-like receptors. To gain insight into the genomic organization, regulation of expression, and supragenomic diversification of GPR34 several vertebrate species were analyzed. In contrast to the obviously intronless coding region GPR34 displays an evolutionary preserved 5' noncoding intron–exon structure. Further, an alternatively used cryptic intron was identified within the coding region, which shortens the N terminus by 47 amino acids. Ubiquitous expression of GPR34 is driven by genomic sequences upstream of at least two transcriptional start regions in mouse and rat but only one region in human. In rodents, both promoters are active in all tissues investigated, but the level of activity is tissue-specific. At the translational level, several conserved in-frame AUGs within the first 150 bp of the coding region may serve as start points for translation in human and other mammals. Combinatory mutagenesis and expression of reporter constructs confirmed these multiple translational start points and revealed a preference for the second in-frame AUG in human GPR34. Our data show that multiple translation initiation starts and alternative splicing contribute to the supragenomic diversification of GPR34.

© 2005 Elsevier Inc. All rights reserved.

**Keywords:** Orphan receptor; GPR34; G-protein-coupled receptor; Genomic organization; Promoter; Translation start

Completion of the human and other vertebrate genome sequencing projects revealed more than 1000 genes belonging to the G-protein-coupled receptor (GPCR) superfamily. Two-thirds of these genes are thought to encode odorant receptors. Of the remaining third, the natural ligand for only half of the receptors has been identified, leaving about 160 so-called “orphan” GPCRs with unknown agonists [1]. Once the sequence of a GPCR is known, understanding the function of the encoded protein becomes a task of paramount importance. Consequently, pharmacological characterization and functional genomic strategies can be employed to identify activating ligands and their functional relevance.

**Abbreviations:** GPCR, G-protein-coupled receptor; hGPR34, human GPR34; mGPR34, mouse GPR34; ORF, open reading frame; ROX, carboxy-X-rhodamine; UTR, untranslated region; RT-PCR, reverse transcription PCR.

<sup>☆</sup> Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. DQ\_106398–DQ\_106402 (human ESTs), DQ\_103764–DQ\_103767 (mouse ESTs), and DQ\_106403 (bonobo GPR34).

<sup>\*</sup> Corresponding author. Fax: +49 341 9722159.

E-mail address: [angela.schulz@medizin.uni-leipzig.de](mailto:angela.schulz@medizin.uni-leipzig.de) (A. Schulz).

In initial studies we and others showed that GPR34 has clear structural and functional features of the group of ADP-like GPCRs [2–5]. This group includes the ADP receptors P2Y<sub>12</sub> and P2Y<sub>13</sub>, the UDP-glucose receptor (GPR105), and the orphan GPCRs GPR87 and GPR34. In mammals GPR34 is located on the X chromosome, making this receptor an interesting candidate gene for human diseases linked to the Xp11.4 locus [6,7].

The presence of GPR34 in cartilaginous and ray-finned fish genomes suggests that the evolution of this receptor likely started in the Cambrian period more than 450 million years ago [3]. While mammalian GPR34 orthologs are single-copy genes encoded by an obviously intronless coding region, a greater GPR34 diversity is found in teleost fishes. Here, two GPR34 genes are observed in many fish species, probably a consequence of genome duplication in early teleost fish evolution [8]. Later, a single intron was introduced into the coding region of only one GPR34 gene. Subsequent gene duplication in some teleost fish species furthered GPR34 diversification [3].

Present knowledge of GPR34 structure is based only on genomic analysis of the coding region in vertebrates. However,

structural diversity can also be determined by allelic variants and supragenomic processes such as splicing and mRNA editing and at the translational level. Therefore, we set out to address these questions from an evolutionary perspective by analyzing allelic variability in the human population, genomic organization, expression regulation, and supragenomic diversification of this orphan GPCR. We found that GPR34 genes have a highly conserved exon–intron structure in mammals that gives rise to splice variants also affecting the GPR34 coding region. Receptor diversification is probably further increased by the presence of multiple translation initiation sites.

## Results and discussion

### *Low allelic diversity of the GPR34 coding region within the human population*

The occurrence of different allelic variants can contribute to GPCR structural and functional diversity. Because GPR34 is an X-chromosomal gene, mutations can become relevant even at low frequency in the population. To search for frequent GPR34 alleles we first analyzed the coding region of 20 randomly chosen human individuals of Caucasian origin. Interestingly, we identified a male hemizygous for a silent GTA to GTG exchange at codon position 296. Analysis of his mother revealed heterozygosity at this position. This initial screen prompted us to extend the analysis. Since genetic diversity is best related to linguistic groups [9], we sequenced the GPR34 coding region in 85 humans representing all major linguistic groups [9]. However, no sequence variations within the GPR34 coding region were found in this worldwide panel. Further, database mining (SNP database at NCBI; <http://www.hapmap.org>) revealed no frequent single nucleotide polymorphisms (SNP) yet identified. Very recently, Iida and Nakamura [10] searched for SNPs in 29 GPCRs, including GPR34, in 48 healthy Japanese individuals and found no frequent SNP in the coding region. This does not exclude individual variations as found in our Caucasian sample set (see above); our observations support the finding that there are no frequent or fixed allelic variants in major human subpopulations.

The chimpanzee (NM\_001009135) and pygmy chimpanzee (bonobo; DQ\_106403) GPR34 sequences differ from their human ortholog at only two silent nucleotide positions. Because chimpanzees showed a significantly higher interindividual sequence diversity compared to humans [11,12], we analyzed the GPR34 coding sequence of 10 chimpanzee individuals. Again, no interindividual sequence differences were found in chimpanzee, indicating that the GPR34 sequence remained highly conserved over the past 5.5 Myr of hominoid speciation. These data support the conclusion that, in contrast to other GPCRs [13], allelic variants do not contribute to the structural diversity of GPR34.

### *GPR34 has a conserved 5'-exon–intron structure*

At the time this study was initiated there was only limited information (EST database) suggesting an exon–intron struc-

ture within the 5' UTR of GPR34. To validate these first hints and to search for heterogeneity of GPR34 transcripts due to alternative splicing or mRNA editing we compared mRNA and genomic sequences of GPR34.

Fifty-two transcripts from human fetal brain and human promyelocytic cell line HL60 cDNA libraries, 36 transcripts of mouse fetal and brain cDNA libraries, and 3 transcripts of a rat spleen cDNA library were analyzed and compared with genomic sequences taken from the database (GenBank Accession Nos. AL627402 (human), AL671117 (mouse), AABR03119244 (rat)). First, we found no evidence of mRNA editing in any of the GPR34 transcripts. Second, in contrast to the discernibly intronless coding sequence, we found an exon–intron structure 5' of the GPR34 coding region (Fig. 1, Table 1). Comparison of the exon–intron structure between human, mouse, and rat genes implicated a homologous transcription start region approximately 6.5 kb upstream of the translation start site. The first two exons are present in all observed human, mouse, and rat GPR34 transcripts (Fig. 1A). In some human and mouse transcripts one or two additional exons were found and could be traced to a genomic sequence positioned between exon 2 and the coding region (see Fig. 1A, rare exons, Table 1). It is of interest to note that the splice acceptor site of human exon 3 is variable. Most commonly used is the acceptor site –79. In several transcripts exon 2 was linked to the –6 splice site (see Fig. 1A, human) but an in-frame stop codon in exon 2 avoids 5' extension of the open reading frame. We also characterized the 3' UTR of human, mouse, and rat GPR34 but found no indications of introns in this region. Most cDNA 5' ends were found in an ~90-bp region upstream of the splice donor site in exon 1 (Figs. 1A and 1B). The high degree of conservation between human and rodent cDNA 5' ends as well as the presence of conserved TATAA and CCAAT boxes in close proximity to this genomic segment are all indicative of a transcriptional start site (further referred to as proximal promoter). Surprisingly, we found several transcripts in the mouse and rat cDNA libraries with a different transcription start point located approximately 700 bp upstream of the main transcriptional start site (Fig. 1, mouse, rat, distal promoter). Here, the region of the main transcriptional start becomes part of an exon (exon 1). All attempts to identify human transcripts that use this distal promoter failed so far. Our data are in good agreement with previous Northern blot studies that identified two main transcripts (2.1 and 2.3 kb) in mouse and rat tissues but not in human samples [2,4]. The sizes of the rodent transcripts we identified by sequencing nicely match with the band sizes found in Northern analysis and it is, therefore, very likely that the 2.1- and 2.3-kb bands represent the products of the proximal and distal promoter, respectively.

By analyzing PCR fragments that had significant differences in length compared to the most frequent transcripts we identified a splice variant of GPR34 lacking 70 bp within the N-terminal coding region. This splicing is due to a cryptic intron within the coding region (see Fig. 1A, human). The splice sites of the cryptic intron in human GPR34 transcripts are also present in the mouse sequence. To date, we have not found a murine Gpr34 transcript in which this cryptic intron is indeed used.

As shown recently, GPR34 is evolutionarily old, being present in all vertebrate classes [3]. Identification of over 60 vertebrate GPR34 orthologs enabled us to analyze the genomic

structure in other mammalian and nonmammalian species. As found in rodents and human, other mammalian GPR34 genes (dog, cow, pig), as well as nonmammalian chicken, *Xenopus*

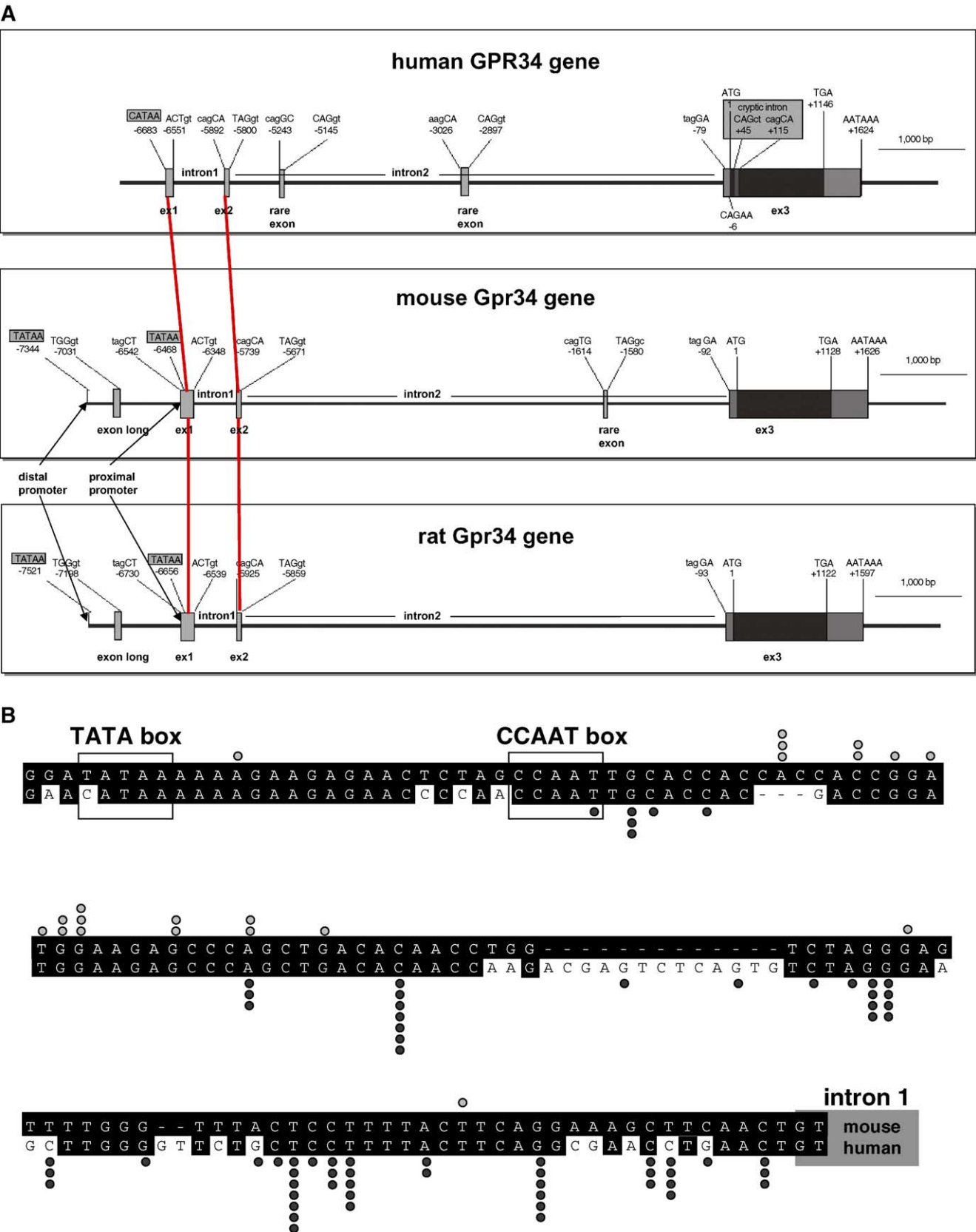


Table 1  
Summary of GPR34 mRNA specimens found in mouse and human

Accession No.	Species	Structure of GPR34 mRNA	Transcript frequency (of all sequenced transcripts)	Transcript frequency (found in EST database)
DQ 106398	Human	Exon 1, exon 2	52/56	56/59
DQ 106399	Human	Exon 1, exon 2, rare exon (–5243/–5145)	1/56	—
DQ 106400	Human	Exon 1, exon 2, rare exon (–3026/–2897)	1/56	1/59
DQ 106401	Human	Exon 1, exon 2, spliced to position –6 instead of –79	1/56	2/59
DQ 106402	Human	Exon 1, exon 2, cryptic intron (+45/+115)	1/56	—
DQ 103764	Mouse	Exon long, exon 1, exon 2	1/36	3/6
DQ 103765	Mouse	Exon long, exon 1, exon 2, rare exon	1/36	—
DQ 103766	Mouse	Exon 1, exon 2	26/36	3/6
DQ 103767	Mouse	Exon 1, exon 2, rare exon	8/36	—

Transcripts from human fetal brain and human promyelocytic cell line HL60 cDNA libraries and from mouse fetal and brain cDNA libraries were cloned and sequenced. Accession numbers of different sequences and exon presence in the transcript and their frequency in our sample set and the EST database are given.

s, and zebrafish GPR34 genes, contain at least one intron 5' of the first in-frame AUG. This indicates that the promoter region was already separated from the GPR34 coding region by an intron very early in vertebrate evolution and that this 5'-intron–exon-structure was maintained in all vertebrate classes yet sequenced.

#### *GPR34 is ubiquitously expressed*

Characterization of the GPR34 gene structure revealed a complex exon–intron structure and at least two putative transcription initiation sites that prompted us to analyze GPR34 expression and expression regulation. Initial Northern blot analysis suggested ubiquitous GPR34 expression in all embryonic and adult tissues investigated [2]. To verify and extend these analyses, PCR was performed using multiple-tissue cDNA libraries from human, mouse, and rat. As shown in Fig. 2A, GPR34 transcripts are expressed in all human and mouse tissues. Further, GPR34 transcripts are also found in HL60 cells and the murine macrophage cell line RAW 264.7 (data not shown).

We next employed rat multiple tissue cDNAs to search for tissue-specific promoter usage. PCR detection of transcripts from both promoters required two setups with different sense primers (Fig. 2B). In the first step, to detect transcripts yielded from the distal promoter, a sense primer was chosen to anneal to the most distant exon (exon long). PCR fragments from this approach contained parts of the most distant exon, exons 1 and 2, and a part of exon 3. The second setup used a sense primer that annealed in exon 1, thereby yielding shorter PCR products than in the previous setup. In all tissues investigated both PCR strategies produced specific fragments, indicating that probably both promoters are active in rodents independent of tissue type (see Materials and methods; data not shown). Finally, we asked

whether promoter usage differs in distinct parts of a chosen organ. Thus, analyses were performed with cDNAs of different heart regions (left and right atrium, ventricle, aortal part) and, as with whole organ, we observed specific fragments for both primer setups (see Fig. 2B).

In order to quantify and verify the activity of both promoters in mice real-time PCR assays were performed for different tissues. As in case of rat Gpr34, parts of the proximal promoter and its transcription start are an integral part of transcripts of the distal promoter (see Fig. 1 and Table 1).  $C_t$  values of transcripts of the distal promoter were determined directly from primer pair 1-derived PCR product. Because parts of the proximal promoter and its transcription start are an integral part of transcripts of the distal promoter (see Fig. 1 and Table 1), primer pair 2 amplifies transcripts yielded from both promoters (represents the total Gpr34 transcript number). Therefore,  $C_t$  values originating from only the proximal promoter had to be determined indirectly by subtraction of the copy number yielded from primer pair 1 from the copy number yielded from primer pair 2. To ensure equal amplification efficiency of both Gpr34 primer pairs standard curves were generated using serial dilutions of cloned Gpr34 cDNA. As shown in Fig. 3,  $C_t$  values for primer pair 1 and primer pair 2 were identical for all mGpr34 cDNA concentrations.

The highest levels of Gpr34 transcripts were found in brain and spleen (Table 2). As found for rat, transcripts from both promoters were found in all murine tissues investigated. Interestingly, activity of the distal promoter was ~14-fold lower in liver compared with the proximal promoter activity in the same tissue (Table 2). In contrast, distal promoter activity was more prominent in heart and spleen. Latter data are in good agreement with Northern blot data showing a more intensive 2.3-kb band, which most likely represents transcripts of the distal promoter (see above).

Fig. 1. GPR34 genomic organization and transcription start determination. (A) To determine the gene structure of GPR34 in human, mouse, and rat, specific mRNAs were analyzed and compared with NCBI database genomic sequences. Exons 1 and 2 (ex1, ex2) are present in all transcripts from the three species (red line). The coding region (black) and splice sites are shown for each exon with numbering referring to the first putative translation start codon. Additional exons located within the second intron region were found in only some of the human and mouse transcripts. Further, a cryptic intron was identified (gray box) within the N-terminal coding region of hGPR34. A second transcriptional start site, upstream of the first, was identified in rodent Gpr34 genes. Transcripts derived from this second promoter region (distal promoter) contained the first transcriptional start as an exon, termed “exon long”. (B) In the majority of human and rodent transcripts, the transcriptional start is located about 6.5 kb upstream of the coding region. Aligned human and rodent sequences upstream of this most frequently used transcription start are shown. These sequences contain conserved TATAA and CCAAT boxes. The number and position of the 5' end of the transcripts are given (encoded by circles: gray, mouse; dark gray, human).



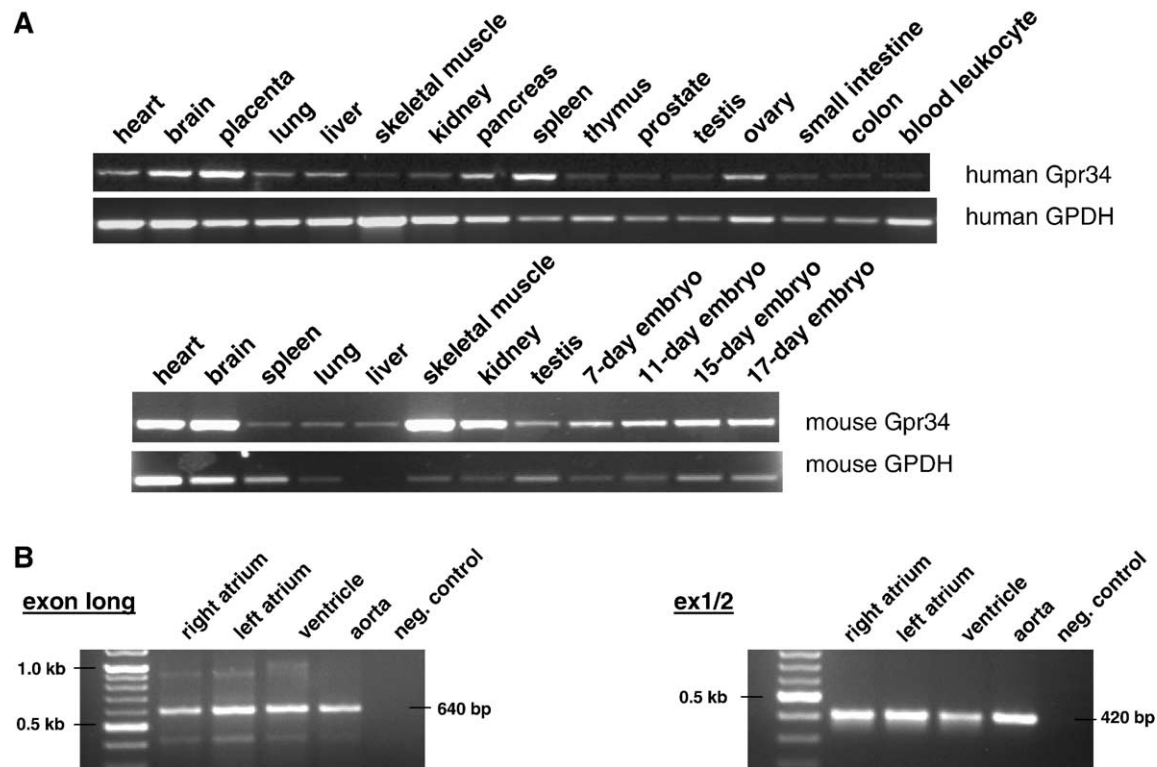


Fig. 2. Ubiquitous expression of GPR34 in different species. The expression pattern of GPR34 was determined by reverse transcriptase PCR. Thus, GPR34-specific primer pairs, located in exon 1 and exon 3 (Fig. 1), and mouse and human cDNA panels were used (see Materials and methods) and (A) PCR results are shown. For cDNA quality control purposes GPDH expression was monitored in parallel. (B) cDNA was prepared from different regions of the rat heart and two sets of primers, located in exon long and exon 3 (left) or in exon 1 and exon 3 (right), were used to monitor the usage of both promoters.

Taken together, our data show that the ubiquitous expression of GPR34 is mediated by at least one promoter in human and two promoters in rodents. As found in mouse there is evidence for tissue-specific preference of one of the two promoter regions, at least in liver.

Promoter studies of GPR34

The ubiquitous expression of GPR34 suggests constitutive and cell-type-independent promoter function. To probe promoter activity further, three mouse genomic fragments, including the conserved transcriptional start; the CCAAT and TATAA boxes; and 0.5, 1.5, and 2 kb of upstream sequences; were cloned into the promoterless pcDNAI/Amp vector (Fig. 4). The firefly luciferase gene was used as reporter. To ensure equivalent mRNA stability and translation, we preserved the sequence environment by cloning the entire mouse Gpr34 5' UTR between the putative promoter sequences and the ATG of luciferase (see Fig. 4). The CMV promoter in the Gpr34 5' UTR context served as a positive control, while constructs without a promoter sequence served as negative controls. COS-7 and RAW cells were transfected with these constructs and luciferase activity was measured. The lower absolute luminescence in RAW cells was due to the significantly lower transfection efficiency in this cell line (0.1% in RAW compared to 43% in COS-7 cells as determined by transfection of GFP-encoding plasmid and counting of fluorescent cells). None of the genomic fragments

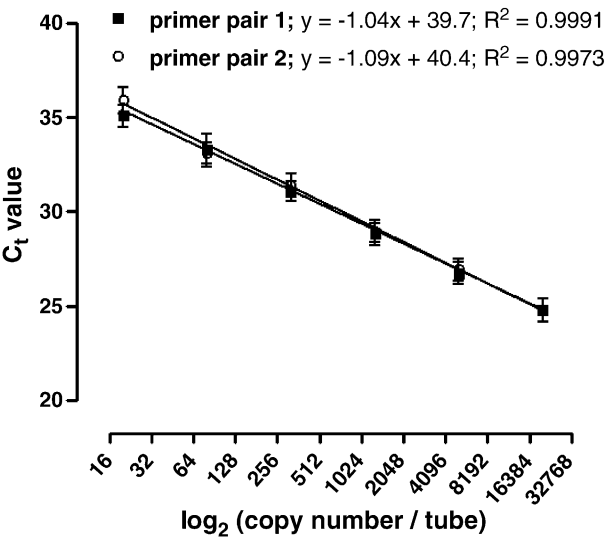


Fig. 3. Real-time PCR standard curves for Gpr34. Real-time PCR was used to quantitate the activity of both Gpr34 promoters in mouse tissue. Primer pair 1 was specific to transcript from the distal promoter. Primer pair 2 amplifies transcripts from both the proximal and the distal promoter because all transcripts of the distal promoter also contained parts of the proximal promoter and its transcription start (see Fig. 1 and Table 1). Therefore, proximal promoter activity can be determined only indirectly by subtraction of the copy number yielded from primer pair 1 from the copy number yielded from primer pair 2 (for equation see Table 2). To ensure equal amplification efficiency of both Gpr34 primer pairs standard curves were generated using serial dilutions of cloned Gpr34 cDNA (Accession No. DQ 103764, see Table 1) containing 20, 80, 320, 1280, 5120, and 20,480 copies per 3  $\mu$ l.

Table 2

Quantification of transcripts from proximal and distal mouse Gpr34 promoter

Tissue	Mouse Gpr34 mRNA			Controls	
	$\Delta C_t^a$ (distal promoter)	$\Delta C_t^a$ (proximal promoter)	Ratio promoter activity <sup>b</sup> (proximal/distal)	$\beta 2$ -Microglobulin ( $C_t$ value)	Clathrin ( $C_t$ value)
Brain	8.56 ± 0.18	8.48 ± 0.39	1.07 ± 0.18	16.5 ± 1.05	16.9 ± 0.74
Liver	15.70 ± 0.42	11.97 ± 1.00	14.37 ± 5.53**	14.7 ± 0.64	17.7 ± 0.74
Heart	10.76 ± 0.6	11.37 ± 0.34	0.67 ± 0.17*	16.4 ± 0.88	18.6 ± 0.92
Spleen	8.39 ± 0.16	9.08 ± 0.56	0.65 ± 0.20*	15.3 ± 1.03	19.0 ± 1.39
Skeletal muscle	11.23 ± 0.93	10.70 ± 0.51	1.60 ± 0.75	17.3 ± 1.69	19.4 ± 1.90
Kidney	12.08 ± 0.30	12.48 ± 0.62	0.88 ± 0.36	15.9 ± 0.86	17.3 ± 1.17

Three micrograms of total RNA from different tissues was subjected to reverse transcriptase reaction. Quantitative real-time PCR (TaqMan assay) was performed for each transcript and controls ( $\beta 2$ -microglobulin, clathrin) as described under Materials and methods.

<sup>a</sup>  $C_t$  values were normalized to the  $C_t$  values of the controls and are given as  $\Delta C_t$  ( $\Delta C_t = C_t - (C_t \beta 2\text{-microglobulin} + C_t \text{clathrin})/2$ ).

<sup>b</sup> The ratio was determined using the equation  $2^{(\Delta C_t \text{ distal} - \Delta C_t \text{ proximal})}$ . Data are presented as means ± SD of four independent experiments (four individual mice) each performed in duplicate. Significant differences between  $C_t$  values of proximal and distal promoter transcripts from the same tissue are marked.

\*  $p = 0.05$ .

\*\*  $p < 0.01$ .

of the putative Gpr34 promoter region showed significant activity in either cell line compared to the expression of promoterless constructs (Table 3).

It is well known that promoter activity can be influenced by downstream intron and genomic sequences [14–19]. Therefore, the intron sequence between exons 1 and 2 (intron 1) was inserted in its genomic context into the original reporter constructs (see Fig. 4, referred to as 0.5-Int, 1.5-Int, and 2.0-Int) and expressed in both cell lines. Again, CMV control vector-transfected cells showed high luciferase activity; however, no specific promoter activity was found for the putative Gpr34 promoter constructs. Proper intron splicing was con-

firmed by sequencing of specific PCR fragments from cDNA of RAW cells transfected with the CMV-Int control vector. It is of interest to note that introduction of the intron significantly increased basal and CMV promoter-driven luciferase activity, probably a consequence of higher mRNA stability, mRNA abundance, or some additional promoter activity of the intron sequence.

To exclude the possibility that the lack of Gpr34 promoter activity was due to context effects caused by the pcDNAI/Amp vector backbone, all intron-containing Gpr34 promoter constructs were subcloned into the pGL3 vector. Here, an SV40 promoter served as positive control. As observed in the

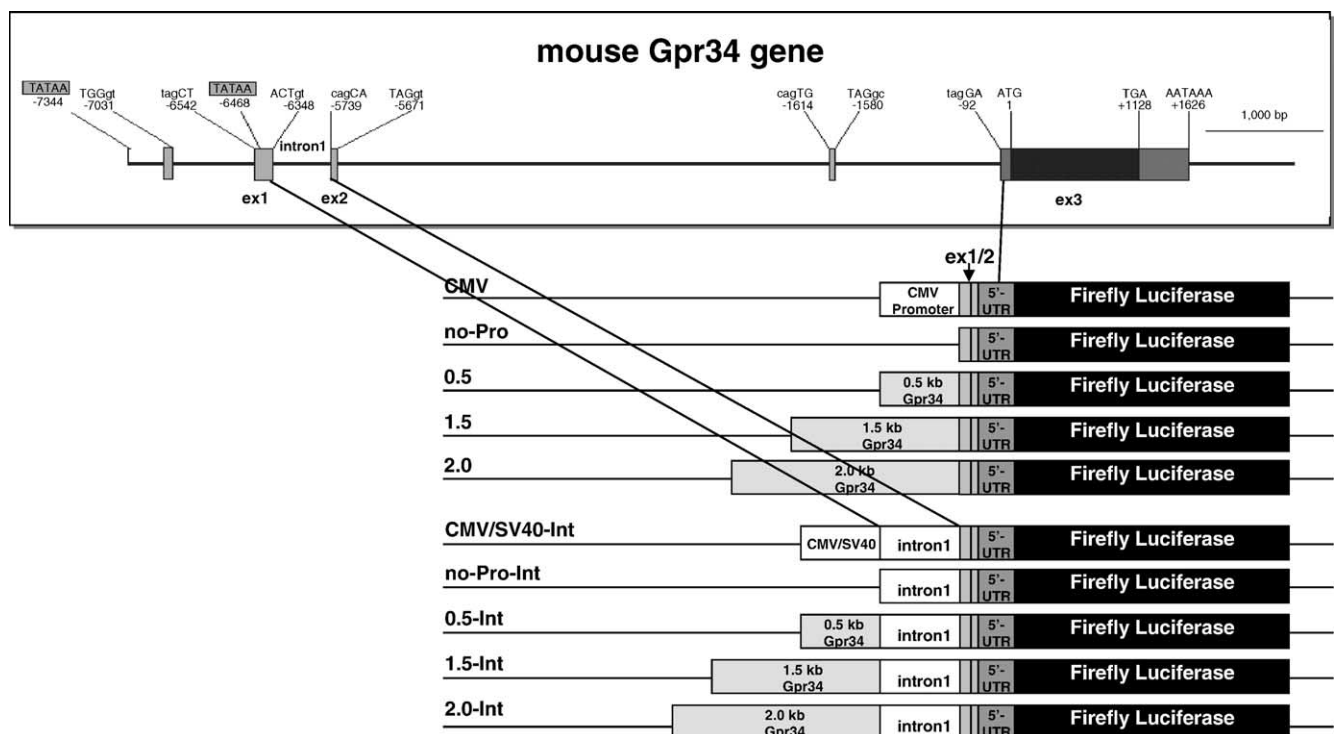


Fig. 4. Promoter constructs. To test for potential promoter activity, 0.5, 1.5, and 2.0 kb of mouse genomic sequence upstream of the major transcriptional start site were cloned into two firefly luciferase-based reporter constructs (see Materials and methods). As a positive control, luciferase expression was driven by CMV and SV40 promoters in pcDNAI/Amp and pGL3 vector, respectively. Additional constructs were generated by cloning intron 1 into the appropriate position within the 5' UTR (see also Table 1).

Table 3  
Luciferase activity of different promoter constructs in two vector systems

Construct	Promoter	Luciferase activity (COS-7 cells) (fold over control)		Luciferase activity (RAW264.7 cells) (fold over control)	
		pcDNAI/Amp	pGL3	pcDNAI/Amp	pGL3
CMV	CMV	55,950 ± 6360	–	42.9 ± 2.0	–
No-Pro	None	593 ± 197	–	6.4 ± 0.2	–
0.5	Gpr34	79 ± 19	–	1.7 ± 0.5	–
1.5	Gpr34	107 ± 30	–	2.4 ± 0.3	–
2.0	Gpr34	139 ± 33	–	3.9 ± 0.3	–
CMV/SV40-Int	CMV/SV40	541,131 ± 105,894	834,564 ± 273,855	634 ± 166.4	63.5 ± 17.5
No-Pro-Int	None	4990 ± 740	888 ± 297	9.3 ± 0.9	5.6 ± 1.6
0.5-Int	Gpr34	8308 ± 2022	563 ± 173	8.6 ± 1.8	6.1 ± 1.4
1.5-Int	Gpr34	1653 ± 617	345 ± 120	6.8 ± 1.2	2.9 ± 0.5
2.0-Int	Gpr34	1685 ± 584	117 ± 38	6.5 ± 1.6	2.6 ± 0.6

For testing the promoter activity of the genomic sequence 5' of the transcription start, different 5' regions of Gpr34 were cloned upstream of the firefly luciferase coding sequence inserted into pcDNAI/Amp and pGL3 vectors as described under Materials and methods. Luciferase activity in COS-7 cells was measured and is presented as fold over negative control (GFP-transfected cells, 24.5 ± 8.6 cps/pcDNAI/Amp; 60 ± 20 cps/pGL3). Results are given as means ± SEM of four (pcDNAI/Amp constructs) or three (pGL3 constructs) independent experiments performed in quadruplicate. For RAW 264.7 cells results for three representative experiments are shown and given as means ± SEM of three (pcDNAI/Amp and pGL3 constructs) independent experiments performed in quadruplicate. Background activity of GFP-transfected cells (negative control) was 27.3 ± 6.1 cps for pcDNAI/Amp and pGL3. cps, counts per second.

pcDNAI/Amp vectors, no specific activity of the putative Gpr34 promoter regions was found in the pGL3 vector constructs.

Finally, we generated a construct containing the Gpr34 coding sequence and 6.7 kb of genomic sequence upstream of the Gpr34 start AUG (0.7 kb putative promoter sequence plus all exons and introns). This minigene construct was transfected into COS-7 cells and expression was monitored by reverse transcribing the mRNA into cDNA followed by PCR and sequencing of PCR products. PCR fragment analysis revealed properly spliced mGPR34 mRNA, indicating that the GPR34 minigene shows promoter activity. However, the high sensitivity of this PCR-based approach was not comparable directly to the results from luciferase reporter assays. Therefore, RT-PCR analysis was repeated for mRNA of 0.5-Int-transfected COS-7 and RAW cells using 5' UTR sense and luciferase-specific antisense primers. As for the minigene-transfected cells, properly spliced cDNA was identified, suggesting some Gpr34 promoter activity. Because these transcripts do not significantly contribute to luciferase activity above control transfected cells (Table 3) we interpret this observation as a leaky transcriptional activity rather than specific promoter-driven expression.

In sum, despite clear identification of the transcriptional start in the rodent and human GPR34 genes, analysis of the upstream sequence provided no clear indications of promoter activity in vitro and, therefore, awaits further studies, e.g., analysis of larger upstream sequences in the context of the used minigene.

#### *Translation of GPR34 is ensured by several in-frame AUG codons*

Translation of mRNA into protein usually starts at the first AUG within the open reading frame (ORF). A scanning mechanism for AUG and distinct sequence patterns that ensure the recognition of this translational start point by the ribosomal machinery has been identified [20]. The scanning mechanism suggests that ribosomes traverse the 5' UTR and that translation

of the major coding domain is reduced when out-of-frame AUG codons occur upstream. Our mRNA analysis revealed that human and murine GPR34 mRNAs usually contain two to six and two to three AUGs, respectively, which are upstream of the first in-frame AUG. The number of upstream AUGs depends on mRNA splicing, and the resulting ORFs are fewer than 22 codons long. If an upstream AUG is used by the ribosomal machinery, then translation of the main ORF requires either reinitiation of translation or a context-dependent leaky scanning [20]. Further, a growing number of studies show that translation start for transcripts with multiple AUG can occur not only at the first AUG within the ORF [21,22]. The presence of several AUGs in the 5' UTR and up to five AUG codons within the first 150 bp of the putative N terminus of mammalian GPR34 (Fig. 5A) raised the question of which of the in-frame AUG codons serves as a translational start point. For example, the second in-frame AUG (Met<sup>8</sup>) in human GPR34 is the first AUG in mouse and rat GPR34 mRNA; thus, we want to determine which in-frame AUG is used in human GPR34. To address this question, the first three in-frame AUG codons were combinatorially mutated to TTG in a hGPR34–GFP fusion construct. As shown in Fig. 5B, all three AUG can initiate robust translation. A significant reduction in GFP fluorescence (55 ± 14.0% of the wild-type construct A1111) was observed when all three AUGs were eliminated (A0001). Similarly, the open reading frame of the hGPR34 splice variant (cryptic intron, see above) lacking 47 amino acids of the N terminus also started with AUG4 and it showed 75 ± 10.4% of the fluorescence of the wildtype construct A1111. This observation is of interest because the transcript of the splice variant still contains the first two AUGs (codons 1 and 8), albeit out of frame. To demonstrate proper hGPR34 expression and localization in the plasma membrane, confocal immunofluorescence studies were performed. All hGPR34–GFP constructs were found in the plasma membrane (data not shown), indicating proper receptor folding and trafficking.

To verify these results and to exclude the possible contribution of translational start from a more downstream AUG, GFP



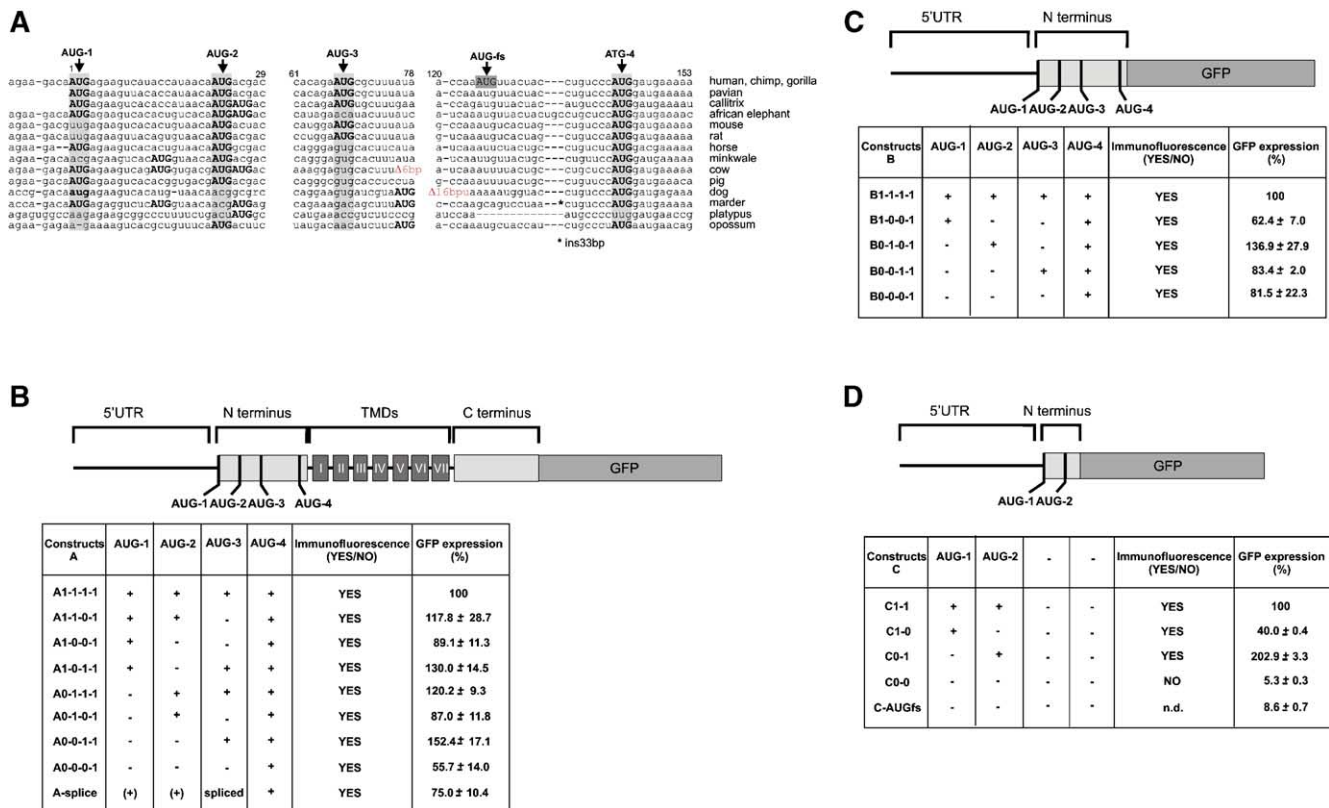


Fig. 5. Translation start analysis in Gpr34. (A) N-termini sequence comparison of different GPR34 orthologs shows multiple in-frame AUGs in Gpr34 and other mammalian GPR34 sequences. (B) To identify the translation initiation codon(s) a combinatory mutagenesis strategy was applied to the first three in-frame AUG codons (AUG to UUG mutation). Translation was monitored by quantification of fluorescence emitted by the various GPR34–GFP fusion constructs (see also C and D). Results are presented as percentage of the corresponding wild-type construct. (C) In a second set of constructs GFP was fused directly to the N terminus of hGPR34. Constructs depicted in panel D contain only the first third of the N-terminal coding region, which was fused to GFP. The putative translation sites at codon positions 1 and 8 were subsequently mutated. Data in the tables are given as means  $\pm$  SEM of three independent experiments, each performed in quadruplicate. n.d.—not determined.

was fused directly to the N-terminal coding region (Fig. 5C). Translation was found for all constructs; however, deletion of the transmembrane coding region left the fusion proteins soluble within the cells as analyzed by confocal microscopy (data not shown). Strikingly, mutation of the second and third AUGs (B1001) resulted in a significant reduction in translation, indicating that AUG2 and AUG3 contribute to maximal translation.

Finally, constructs containing only the first two potential translation AUGs were tested. As shown in Fig. 5D, either AUG was capable of translation initiation, whereas mutation of both AUG codons resulted in a loss of expression. As found in B constructs, the second AUG significantly contributes to translation and appears to be most efficiently used, although the context ( $-3$  to  $-1$  and  $+4$  position) of both AUGs is identical in the human sequence. GPR34 may therefore represent a new example that demonstrates escape from the first-AUG rule [20].

Our experiments showed that translation can occur from all four in-frame AUG codons within the human GPR34 N terminus when each of these AUGs is set first in the ORF. It is well established that both the AUG and its sequence context determine the efficacy of translation start initiation [20]. One may argue that AUG codon usage by the translation apparatus

lacks specificity in our heterologous expression system since each AUG is used to some extent, but cells transfected with a construct bearing an endogenous out-of-frame AUG brought in-frame to GFP (AUGfs) showed no fluorescence above control (GFP without AUG) (Figs. 5A and 5D). These findings confirm the specificity of expression driven by endogenous in-frame AUG codons.

Taken together, the 5' UTR and the ORF of several mammalian GPR34 contain more than one potential translation start AUG and, at least in hGPR34, there is evidence of alternative usage, most likely through a leaky-scanning mechanism. Multiple AUG codons at the very beginning of the ORF in close proximity may have an advantage to ensure and maximize translation. Furthermore, this gene structure guarantees translation even if one AUG got lost by mutation or during evolution. The multiple translation starts may also contribute to supragenomic diversity in receptor length. This is supported by previous Western blot studies of hGPR34 [24]. Here, the specific immunostaining indicates reasonable molecular weight diversity, and receptor length differences may contribute to these findings. One can speculate that different “receptor isoforms” may have different functions as shown for other proteins [23,25]. Unless the natural agonist is identified this question remains to be addressed.

## Conclusion

GPR34 displays remarkably low allelic diversity over the past 5 Myr of hominoid evolution and allelic variants are rare in the human population. The 5' UTR of GPR34 transcripts is derived from a conserved intron–exon structure upstream of the coding region. Despite aberrant splicing of cryptic introns within both this region and the N-terminal coding region, the gene structure does not contribute significantly to generating isoforms of this single-copy gene. However, several AUG codons in the 5' UTR and at the very beginning of the main ORF suggest alternative initiation of translation. Indeed, several lines of experimental evidence implicate a remarkable leaky scanning of GPR34 mRNA for start AUG codons, which may provide the basis of length isoforms of GPR34.

## Materials and methods

### Genomic sequence analysis of the human GPR34

To identify individual variations we sequenced the GPR34 coding region in 105 humans (85 individuals of the DNA panel as described previously [9] and 20 German individuals). PCR fragments were amplified with genomic DNA samples and the primer pair 5'-CACTTTTGTATCCACAGAAGACA-3'/5'-GTTCTTCAAAGAATTAAGAGGCTGC-3'. Genomic DNA (100 ng) was used in PCRs (50 µl) with primers (20 pmol each), standard buffer (Perkin–Elmer Applied Biosystems, Foster City, CA, USA), dNTP (200 µM), and *Taq* polymerase (1 U; Perkin–Elmer). The reactions were initiated with a denaturation at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 45 s, annealing at 60°C for 45 s, and elongation at 72°C for 2 min. A final extension step was performed at 72°C for 10 min. PCR products were separated in a 1.5% agarose gel, purified using a PCR Product Purification Kit (Qiagen), and sequenced.

### Generation of cDNA libraries from different tissues and cell lines

Large-scale GPR34 mRNA analysis via rapid amplification of cDNA ends (RACE) of human, mouse, and rat tissue was utilized for comparative analysis of the genomic structure and identification of editing events, splice variants, and transcription start sites. cDNA was generated from total RNA from the human promyelocytic cell line HL60 (which expresses GPR34) and rat spleen using Fenzol according to the manufacturer's protocol (Active Motif; Rixensart, Belgium). Total RNA was transcribed into cDNA libraries using the GeneRacer Kit (Invitrogen, Karlsruhe, Germany). The amplification of GPR34 cDNAs was performed using specific antisense primers (human, 5'-AAATACATAGAGGGCGATTATGTTCCCAACCAG-3'; mouse and rat, 5'-CAGAAATACATAGAGGGCAATGATGTTTCCAACCA-3') and RACE primers provided with the kit. Additionally, RACE analysis was performed with the human fetal brain Marathon-Ready cDNA library according to the manufacturer's protocol (BD Biosciences, Palo Alto, CA, USA). For analysis of mouse Gpr34 transcripts we used a brain and fetal mouse Marathon-Ready cDNA library from leukocytes (BD Biosciences) with the gene-specific antisense primer 5'-CAGAAATACATAGAGGGCAATGATGTTTCCAACCA-3' and the cDNA library linker primer AP1.

PCR products from all libraries were cloned into the pCR2.1-TOPO vector (Invitrogen) and sequenced with Thermosequenase and dye-labeled terminator chemistry by an automated sequencer (Applied Biosystems, Foster City, CA, USA).

### Expression pattern analysis of GPR34

For sensitive characterization of GPR34 expression in mouse and human tissues, PCR analyses were performed using cDNA mouse panel I and human panels I and II (BD Biosciences). To exclude amplification of genomic

contaminations within cDNA libraries primers were complementary to GPR34 exon sequences that flank introns in the genomic sequence. For mouse PCR the primer pair 5'-TTGGGTTTACTCCTTTTACTTCAGG-3'/5'-CAGAAATACATAGAGGGCAATGATGTTTCCAACCA-3' and for human cDNA the primer pair 5'-GGTCTGCTCCTTTTACTTCAGGCG-3'/5'-AAATACATAGAGGGCGATTATGTTCCCAACCAG-3' were used. As control GPDH fragments were amplified with primers provided by the manufacturer (BD Biosciences).

For expression analysis in rat, tissues were dissected (brain, heart, aorta, spleen, skeletal muscle, esophagus, stomach, small intestine, appendix, rectum, liver, bladder, kidney, suprarenal gland, testes, epididymides) and total RNA was prepared as described above. Reverse transcription was performed using a mixture of oligo(dT) primer and random hexamers. The resulting cDNAs were used in PCRs with the following primer sets: 5'-TTTTCCTTGAAATAGGAGAGAGCAC-3'/5'-CAGAAATACATAGAGGGCAATGATGTTTCCAACCA-3' (transcript exon long) and 5'-TTGGGTTTCTCTCTTACTTCAGC-3'/5'-CAGAAATACATAGAGGGCAATGATGTTTCCAACCA-3' (transcript exon 1/2, Fig. 2B).

### Real-time PCR assay

Total RNA was isolated from approximately 20 mg of tissues using Trizol (Invitrogen) according to the manufacturer's protocol. Complementary DNA was synthesized from 3 µg of total RNA with reverse transcriptase (Superscript II kit; Invitrogen) and an oligo(dT) primer. Quantification of gene expression was performed by real-time PCR (ABI Prism 7700 sequence detection system (Perkin–Elmer)) and SybrGreen (Invitrogen) according to the manufacturer's protocol. For each reaction 25 µl of SybrGreen mix, 3 µl of cDNA template solution, 2 × 1.5 µl of primer (10 µM), 1 µl of ROX, and 18 µl of H<sub>2</sub>O were used. To quantify murine GPR34 transcripts two primer pairs were used: pair 1, 5'-GCCTCAGCATCTTTTCCTTG-3'/5'-GGTACAGTTGTTTCCCTTTT-3', and pair 2, 5'-AACTCTAGCCAATTGCACCACC-3'/5'-GGCCATACTTCAGAAGCTGCA-3'. The point at which the amplification plot crossed the threshold was defined as  $C_t$ , which represented the cycle number at this point. The  $C_t$  values of transcripts of the distal promoter were determined directly from primer pair 1-derived PCR product.  $C_t$  values originating only from the proximal promoter had to be determined indirectly by subtraction of the copy number yielded from primer pair 1 from the copy number yielded from primer pair 2. In all TaqMan assays the transcript levels of the housekeeping proteins clathrin (primer pair 3, 5'-ACAGAGACACAGCCCATGTT-3'/5'-CCCA-TACGGTGGTGCAGTAT-3') and  $\beta$ 2-microglobulin (primer pair 4, 5'-GCTATCCGAAACCCCTCAA-3'/5'-GGCGGGTGGAACTGTGTTA-3') were determined and used to normalize data of each tissue between mouse individuals.  $C_t$  values were normalized to the  $C_t$  values of the controls and given as  $\Delta C_t$  ( $\Delta C_t = C_t - (C_t \beta\text{-microglobulin} + C_t \text{clathrin})/2$ ).

### Generation and luciferase activity assay of Gpr34-promoter constructs

To analyze functionally potential promoter sequences of mGpr34, firefly luciferase was amplified from the pUHC-13-3 vector (BD Biosciences) and cloned into the pcDNA1-Amp vector (referred to as Luc-pcDNA1/Amp). To ensure an equal sequence context of the mRNA upstream of the putative translational start, the entire 5' UTR sequence of the murine mRNA, including positions –1 to –95, exon 2 (–5673 to –5739), and exon 1 (–6347 to –6355) (see Fig. 1), was amplified from cDNA prepared from the murine macrophage cell line RAW 264.7 and inserted 5' of the start ATG of luciferase (referred to as no-Pro; see Fig. 4). To test whether intron 1 (located between exons 1 and 2, Fig. 1A) of mGPR34 has an influence on promoter activity, the sequence of intron 1 was inserted between exon 1 and exon 2 (Fig. 4, referred to as no-Pro-Int). On the basis of these two constructs (no-Pro and no-Pro-Int) 0.5 kb (positions –6359 to –6821), 1.5 kb (positions –6359 to –7816), and 2.0 kb (positions –6359 to –8289) of the putative promoter sequence were inserted upstream of the 5' UTR sequence (see Fig. 4). For control purposes, the CMV promoter sequence was inserted into both basic constructs, yielding constructs CMV and CMV-Int (see Fig. 4).

To verify the results with pcDNA1-Amp vector constructs, the pGL3 vector (Promega, Mannheim, Germany) was used as plasmid backbone. Here, the

mouse 5' UTR and intron 1 sequences were inserted between the putative promoter regions (0.5, 1.5, 2.0 kb) and the start ATG of luciferase (see Fig. 4). Basal luciferase activity was determined with a construct not containing the putative promoter regions and a construct containing the SV40 promoter served as positive control. The sequence of all constructs was verified by restriction analysis and sequencing of PCR-derived fragments.

The luciferase activity assay was performed with a luciferase activity detection system based on the method described in van Leeuwen et al. [26]. Briefly, COS-7 cells and RAW 264.7 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 100 U/ml penicillin, and 100 µg/ml streptomycin at 37°C in a humidified 7% CO<sub>2</sub> incubator. COS-7 cells were seeded in 12-well plates (180,000 cells per well) and transfected with 5 µg of plasmid DNA per well by a Ca<sup>2+</sup>-coprecipitation method [27]. Transfection of RAW cells was performed with 5 µl of Rotifectene (Roth, Karlsruhe, Germany) and 1 µg of plasmid DNA per well in a 12-well plate. Medium was changed 3 h later and 2 days after transfection cells were washed with PBS. Cell lysis was induced by adding 100 µl of lysis buffer (77 mM K<sub>2</sub>HPO<sub>4</sub>, 23 mM KH<sub>2</sub>PO<sub>4</sub>, 0.2% Triton X-100, 1 mM DTT, pH 5.8). Twenty microliters of lysate was mixed with 100 µl of luciferase buffer (20 mM Tricine, 2.67 mM MgSO<sub>4</sub>, 0.1 mM EDTA, 33.3 mM DTT, 270 µM coenzyme A, 530 µM ATP, 470 µM D-luciferin) and luminescence was measured for 1 s in the Victor 2-1420 Multilabel counter (Wallac–Perkin Elmer LAS, Jügesheim, Germany).

#### Analysis of translational start points in human GPR34

To identify the translation initiation codon(s) a combinatory mutagenesis strategy of the first three in-frame AUG codons of hGPR34 was performed. Translation was monitored by measurement of the fluorescence emitted by GPR34–GFP fusion constructs generated by fusing the coding sequence of GFP (without the translation initiation AUG) to the C terminus of hGPR34 without the endogenous stop codon by an overlap PCR strategy. To simulate the natural sequence environment in the pcDps expression vector [28], a 108-bp fragment of the hGPR34 5' UTR region (positions –108 to –1) was inserted directly upstream of the first in-frame AUG. The first three in-frame AUG codons were next combinatorially changed by mutating AUG to UUG (construct series A, Fig. 5B).

In the second set of constructs (construct series B, Fig. 5C), only the coding region of the N terminus was fused to GFP, and AUG codons were mutated. For control purposes an out-of-frame AUG codon located at position +125 in the coding region was set in-frame to GFP.

The last constructs (construct series C, Fig. 5D) were further shortened, such that only codons 1 to 15 were maintained, and fused to GFP and contained only the first and the second in-frame AUG. The identities of the various constructs and the correctness of all PCR-derived sequences were confirmed by restriction analysis and sequencing.

Immunofluorescence studies were carried out to examine the subcellular distribution of the different AUG–GPR34/GFP chimeras. COS-7 cells were seeded into six-well plates containing sterilized glass coverslips and transfected as described above. Approximately 48 h later, cells were fixed and fluorescence images were obtained with a confocal laser-scanning microscope (LSM 510; Carl Zeiss Jena, Jena, Germany).

For quantification of GFP expression, COS-7 cells were seeded and transfected as described for the luciferase assay (see above). Two days after transfection cells were washed three times with PBS and then detached by adding 450 µl of 5 mM EDTA in PBS. Measurement of GFP fluorescence was performed with 200-µl samples in the Victor 2-1420 multilabel counter (Wallac) for 0.1 s, using a 460 nm excitation filter and 525 nm emission filter.

#### Acknowledgments

We thank Çigdem Çetindag for excellent technical assistance. We are grateful to Hillary Creely (MPI for Evolutionary Anthropology, Department of Evolutionary Genetics, Leipzig, Germany) and the anonymous reviewers for suggestions and critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft.

#### References

- [1] S.M. Foord, Receptor classification: post genome, *Curr. Opin. Pharmacol.* 2 (2002) 561–566.
- [2] T. Schöneberg, A. Schulz, R. Grosse, R. Schade, P. Henklein, G. Schultz, T. Gudermann, A novel subgroup of class I G-protein-coupled receptors, *Biochim. Biophys. Acta* 1446 (1999) 57–70.
- [3] A. Schulz, T. Schöneberg, The structural evolution of a P2Y-like G-protein-coupled receptor, *J. Biol. Chem.* 278 (2003) 35531–35541.
- [4] A. Marchese, M. Sawzdargo, T. Nguyen, R. Cheng, H.H. Heng, T. Nowak, D.S. Im, K.R. Lynch, S.R. George, B.F. O'Dowd, Discovery of three novel orphan G-protein-coupled receptors, *Genomics* 56 (1999) 12–21.
- [5] P. Joost, A. Methner, Phylogenetic analysis of 277 human G-protein-coupled receptors as a tool for the prediction of orphan receptor ligands, *Genome Biol.* 3 (2002) 63.
- [6] C.A. Boucher, C.A. Sargent, T. Ogata, N.A. Affara, Breakpoint analysis of Turner patients with partial Xp deletions: implications for the lymphoedema gene location, *J. Med. Genet.* 38 (2001) 591–598.
- [7] F.K. Jacobi, M. Broghammer, K. Pesch, E. Zrenner, W. Berger, A. Meindl, C.M. Pusch, Physical mapping and exclusion of GPR34 as the causative gene for congenital stationary night blindness type 1, *Hum. Genet.* 107 (2000) 89–91.
- [8] A. Christoffels, E.G. Koh, J.M. Chia, S. Brenner, S. Aparicio, B. Venkatesh, Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes, *Mol. Biol. Evol.* 21 (2004) 1146–1151.
- [9] H. Kaessmann, F. Heissig, A. von Haeseler, S. Pääbo, DNA sequence variation in a non-coding region of low recombination on the human X chromosome, *Nat. Genet.* 22 (1999) 78–81.
- [10] A. Iida, Y. Nakamura, Identification of 156 novel SNPs in 29 genes encoding G-protein coupled receptors, *J. Hum. Genet.* 50 (2005) 182–191.
- [11] A. Fischer, V. Wiebe, S. Pääbo, M. Przeworski, Evidence for a complex demographic history of chimpanzees, *Mol. Biol. Evol.* 21 (2004) 799–808.
- [12] H. Römpler, A. Schulz, C. Pitra, G. Coop, M. Przeworski, S. Pääbo, T. Schöneberg, The rise and fall of the chemoattractant receptor GPR33, *J. Biol. Chem.* 280 (2005) 31068–31075.
- [13] W. Sadec, E. Hoeg, J. Lucas, D. Wang, Genetic variations in human G protein-coupled receptors: implications for drug therapy, *AAPS PharmSci.* 3 (2001) E22.
- [14] A.R. Brooks, B.D. Blackhart, K. Haubold, B. Levy-Wilson, Characterization of tissue-specific enhancer elements in the second intron of the human apolipoprotein B gene, *J. Biol. Chem.* 266 (1991) 7848–7859.
- [15] M. Yamada, M. Yoshida, K. Hashinaka, Identification of transcriptional cis-elements in introns 7 and 9 of the myeloperoxidase gene, *J. Biol. Chem.* 268 (1993) 13479–13485.
- [16] B.P. Herrmann, L.L. Heckert, Silencing of FSHR occurs through a conserved, hypersensitive site in the first intron, *Mol. Endocrinol.* (2005) 2112–2131.
- [17] M.L. Hastings, H.A. Ingle, M.A. Lazar, S.H. Munroe, Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA, *J. Biol. Chem.* 275 (2000) 11507–11513.
- [18] M.P. Stemmler, A. Hecht, R. Kemler, E-cadherin intron 2 contains cis-regulatory elements essential for gene expression, *Development* 132 (2005) 965–976.
- [19] A. Kappel, V. Ronicke, A. Damert, I. Flamme, W. Risau, G. Breier, Identification of vascular endothelial growth factor (VEGF) receptor-2 (Flk-1) promoter/enhancer sequences sufficient for angioblast and endothelial cell-specific transcription in transgenic mice, *Blood* 93 (1999) 4284–4292.
- [20] M. Kozak, Pushing the limits of the scanning mechanism for initiation of translation, *Gene* 299 (2002) 1–34.
- [21] D.J. Mancuso, C.M. Jenkins, R.W. Gross, The genomic organization, complete mRNA sequence, cloning, and expression of a novel human intracellular membrane-associated calcium-independent phospholipase A(2), *J. Biol. Chem.* 275 (2000) 9937–9945.

- [22] L. Pascual-Le Tallec, C. Demange, M. Lombes, Human mineralocorticoid receptor A and B protein forms produced by alternative translation sites display different transcriptional activities, *Eur. J. Endocrinol.* 150 (2004) 585–590.
- [23] G.D. Spotts, S.V. Patel, Q. Xiao, S.R. Hann, Identification of downstream-initiated c-myc proteins which are dominant-negative inhibitors of transactivation by full-length c-myc proteins, *Mol. Cell. Biol.* 17 (1997) 1459–1468.
- [24] T. Schöneberg, A. Schulz, H. Biebermann, T. Hermsdorf, H. Römpler, K. Sangkuhl, Mutant G-protein-coupled receptors as a cause of human diseases, *Pharmacol. Ther.* 104 (2004) 173–206.
- [25] C.F. Calkhoven, C. Müller, A. Leutz, Translational control of C/EBPalpha and C/EBPbeta isoform expression, *Genes Dev.* 14 (2000) 1920–1932.
- [26] W. van Leeuwen, M.J.M. Hagendoorn, T. Ruttnik, The use of the luciferase reporter system for in planta gene expression studies, *Plant Mol. Biol. Rep.* 18 (2000) 143a–143t.
- [27] M. Hitt, A.J. Bett, C.L. Addison, L. Prevec, F.L. Graham, Techniques for human adenovirus vector construction and characterization, in: K.W. Adolph (Ed.), *Viral Gene Techniques*, Academic Press, San Diego, 1995, pp. 13–30.
- [28] T. Schöneberg, J. Liu, J. Wess, Plasma membrane localization and functional rescue of truncated forms of a G protein-coupled receptor, *J. Biol. Chem.* 270 (1995) 18000–18006.



## 4 Ergänzungen zur vorgenannten Originalarbeit

### 4.1 Vergleich der Expression von GPR34 und GPR34-like Rezeptoren im Huhn

Wie bereits in Northern Blot Studien von Marchese und Schöneberg<sup>17,18</sup> nachgewiesen werden konnte, ist der Säuger-GPR34 in allen untersuchten Geweben exprimiert. Um ein Profil der GPR34-Expression eines Vertebraten, der nicht zur Klasse der Säugetiere gehört, zu erhalten, wurde die Menge der mRNA des Rezeptors in verschiedenen Geweben des Huhns untersucht. Parallel dazu wurde der im Huhn, aber nicht in Säugern vorkommende GPR34-like-Rezeptor analysiert. Es zeigte sich bei beiden Rezeptoren eine ubiquitäre Expression, die mittels der Real-Time-PCR quantifiziert werden konnte. Generell lässt sich feststellen, dass, geht man von einer gleichen Amplifikationseffizienz der PCR-Ansätze für die beiden Rezeptor-mRNAs aus, der GPR34 in den Organen stärker exprimiert wird als der GPR34-like.

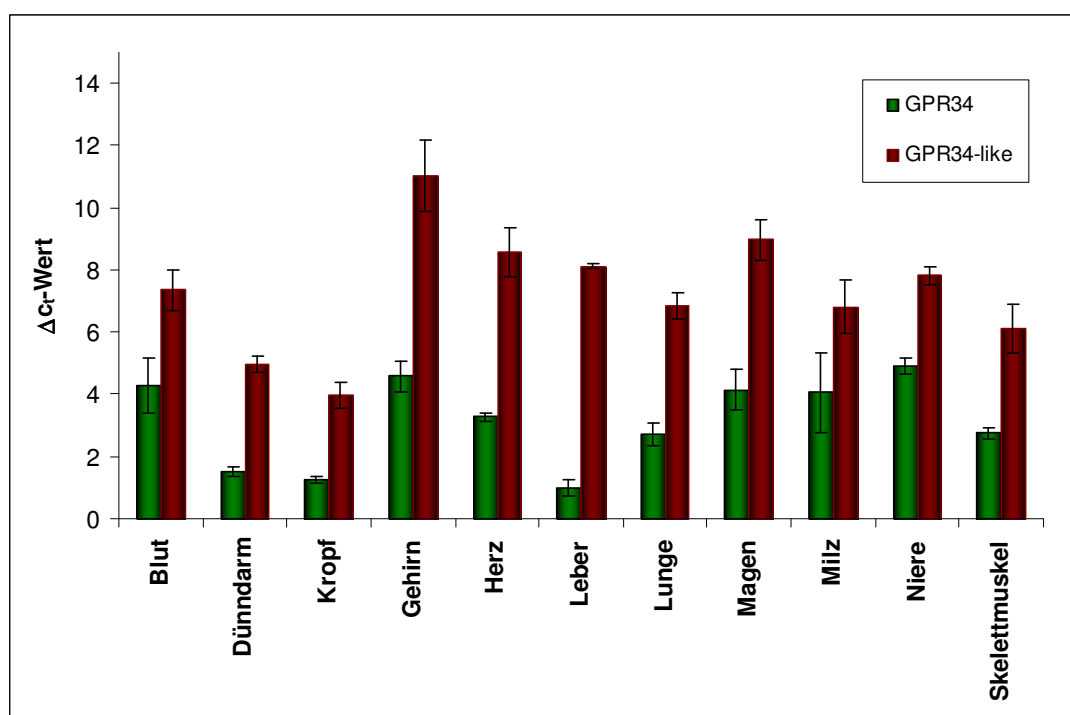


Abbildung 4: Quantifizierung der mRNA Transkripte von GPR34 und GPR34-like in Relation zu Clathrin im Haushuhn (*Gallus gallus*). Dargestellt sind die organspezifischen  $\Delta c_t$ -Werte für den GPR34 (= Differenz von  $c_t$ -Wert GPR34 und  $c_t$ -Wert Clathrin) und GPR34-like (= Differenz von  $c_t$ -Wert GPR34-like und  $c_t$ -Wert Clathrin) in einem Real-Time-PCR Ansatz mit 3  $\mu$ g total RNA. Die Bestimmung der einzelnen  $c_t$ -Werte erfolgte in Doppelbestimmung aus drei unabhängigen Experimenten (drei Versuchstiere) in elf Organen bzw. Geweben mit folgenden qPCR-Primern: Huhn GPR34 Primerpaar: 5'-CCCTGTCTTCAGCAGTGGTT-3' und 5'-GGTCAGTAAGTCAGCTGAAGTGG-3'. Huhn GPR34-like Primerpaar: 5'-CATCGTGCTGCTTCTCTTCC-3' und 5'-TCTCGTGCTGGTCCTCTTGT-3', Huhn Clathrin Primerpaar: 5'-GCTGGAAGCAGAGTGTGGAG-3' und 5'-CTGCAAGAACCACTGCAACA-3'.

Die durchschnittliche Differenz zwischen beiden  $\Delta c_t$ -Werten betrug ca. 3 bis 4, was einem Verhältnis von GPR34 zu GPR34-like mRNA von 8:1 bzw. 16:1 entspricht. Diese Differenz

ist allerdings in Hirn, Herz und Leber höher, was teilweise auf eine höhere GPR34-, aber auch auf eine verringerte GPR34-like-Expression zurückzuführen ist. Eine organspezifische Expression des GPR34-like lässt sich aus den Daten nicht ableiten.

## **4.2 Populationsgenetische Untersuchungen des humanen GPR34**

Um die genetische Variabilität des GPR34 in der menschlichen Population zu untersuchen, wurden zum ersten 20 zufällig ausgewählte Europäer in der GPR34-kodierenden Sequenz untersucht. Dabei konnte ein männliches Individuum mit einem A zu G Austausch in der DNA Sequenz an Position 296 festgestellt werden. Diese Mutation führt zu keiner Veränderung in der Aminosäuresequenz (stille Mutation). Die ebenfalls untersuchte Mutter des Individuums wies an dieser Position eine Heterozygotie auf. Nachfolgend wurden dDNA-Proben von 85 Individuen aus der ganzen Welt untersucht. Die Verteilung ist in Abbildung 5 dargestellt. Von allen 85 Individuen wurde die kodierende Sequenz des GPR34 sequenziert. Allerdings wies keines der 85 Individuen einen Polymorphismus im GPR34 auf.

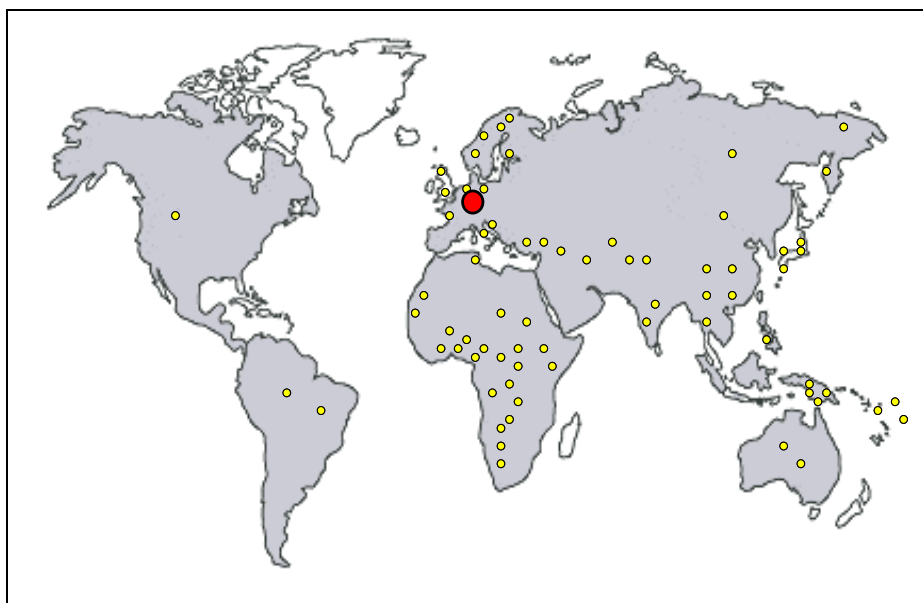


Abbildung 5: Globale Verbreitung der Individuen des untersuchten DNA-Panels. Im Bild gelb markiert sind einzelne oder mehrere untersuchte Individuen einer Region, rot markiert ist die Lokalisation des Individuums mit einem Nukleotidaustausch im kodierenden Bereich des GPR34. Dieser einzelne Nukleotid-Polymorphismus hat keine Auswirkungen auf die Aminosäuresequenz des Rezeptors.

Die GPR34-Orthologe in Schimpansen und Bonobos unterscheiden sich vom Menschen in nur zwei Nukleotidpositionen. Auch diese Austausche sind stille Mutationen und führen somit zu keiner Aminosäureveränderung im GPR34.

## **5 Zusammenfassung der Arbeit**

Dissertation zur Erlangung des akademischen Grades

Dr. rer. med.

Die strukturelle und funktionelle Evolution des G-Protein-gekoppelten Rezeptors GPR34

eingereicht von: Eva Engemaier

angefertigt am Institut für Molekulare Biochemie der Medizinischen Fakultät Leipzig

betreut von Prof. Dr. med. T. Schöneberg

September 2011

### **5.1 Spezifische Promotoraktivität in verschiedenen Organen der Maus**

Bei der Analyse der humanen und murinen GPR34 mRNA-Transkripte konnten ein (humaner) bzw. zwei (murine) putative Transkriptionsstarts ca. 6,5 kbp upstream vom Translationsstart identifiziert werden. In der Maus wird die GPR34-Expression durch zwei Promotorregionen reguliert.

Die durch die beiden murinen Promotoren (als distal und proximal, bezogen auf den Transkriptionsstart, bezeichnet) initiierten Transkripte wurden in einer quantitativen Real-Time-PCR näher charakterisiert. Dabei konnte festgestellt werden, dass eine spezifische Promotoraktivität in bestimmten Geweben nachweisbar ist. Beispielsweise waren in der Leber 14-mal mehr Transkripte des proximalen Promotors als Transkripte des distalen Promotors vorhanden. In Herz und Milz der Maus überwog hingegen die Aktivität des distalen Promotors die des proximalen um circa das 1,5-fache. In den anderen untersuchten Geweben (Großhirn, Skelettmuskel, Niere) waren die Aktivitäten der beiden Promotoren auf ungefähr dem gleichen Niveau. Aus diesen Ergebnissen lässt sich ableiten, dass eine spezifische Aktivität der beiden Promotoren in bestimmten Geweben erforderlich ist. Möglicherweise

werden die daraus entstehenden Transkripte unterschiedlich abgebaut bzw. verfügen über unterschiedliche Signalsequenzen. Dies könnte eine unterschiedliche Regulation des Rezeptors auf mRNA-Ebene bedingen.

Diese putativen Promotorsequenzen wurden anschließend in einem Luciferase-Reportergenassay (Vektorhintergrund pcDNAI/Amp) auf ihre Promotoraktivität getestet. Sowohl in COS-7 als auch in RAW246.7-Zellen ließ sich mit den Konstrukten keine Promoteraktivität im Vergleich zum promotorlosen Konstrukt nachweisen. Dabei war die geringere absolute Luciferaseaktivität in den RAW246.7-Zellen der geringeren Transfektionseffizienz dieser Zelllinie geschuldet. Durch das Einfügen des Introns zwischen Exon 1 und Exon 2 stieg zwar die basale und durch den CMV-Promoter initiierte Luciferaseaktivität an, aber auch hier konnte keine spezifische Promotoraktivität nachgewiesen werden. Diese Steigerung der basalen Luciferaseaktivität könnte durch eine höhere Stabilität der mRNA oder eine leichte Promoteraktivität des Introns bedingt sein. Um einen Einfluss des Vektorgrundgerüsts pcDNAI/Amp auf die Promotoraktivität auszuschließen, wurden alle mGPR34-Promotor-, 5'UTR- und Intron-enthaltenden mGPR34 Sequenzabschnitte in den Vektor pGL3 kloniert. Auch hier konnte nach Transfektion der COS-7 bzw. RAW 246.7 Zellen keine spezifische Promotoraktivität der GPR34-Promoterfragmente nachgewiesen werden.

Ein korrektes Spleißen der mRNA konnte durch eine Transkriptanalyse der mit dem CMV-Promotorkonstrukt transfizierten Zellen sichergestellt werden, so dass alle Komponenten, die für eine posttranskriptionelle Prozessierung nötig sind, auch wirklich im experimentellen Setup vorhanden waren. Möglicherweise liegen nicht alle notwendigen Promotorsequenzen im untersuchten genomischen Bereich. Weitergehende Untersuchungen wären darauf aufbauend erforderlich.

## **5.2 Alternative Translationsstarts im hGPR34**

Die Translation von mRNA zu Protein beginnt im Normalfall am ersten AUG des 5'-Bereiches der mRNA. Dafür wurden ein AUG-Scanmechanismus und bestimmte Sequenzmotive (Kozak-Sequenz) postuliert, die sicherstellen, dass das Ribosom den Translationsstart erkennt<sup>25</sup>. Die im Vergleich der N-Termini verschiedener GPR34-Orthologe festgestellten, mehrfach vorhandenen Methionine sind in der Nukleotidsequenz als AUG codiert und damit als möglicherweise ebenfalls als Startcodon nutzbar. Auf der Basis des humanen GPR34 fusioniert mit GFP wurden drei Konstruktsreien mit kombinatorisch mutierten AUG erstellt, um zu überprüfen, ob diese AUGs auch tatsächlich als



Translationsstart durch die Ribosomen erkannt und verwendet werden. In diesen Experimenten konnte festgestellt werden, dass alle vier im N-Terminus-kodierenden Bereich des humanen GPR34 vorkommenden AUG, die im Leserahmen lagen, als Translationsstart genutzt werden können. Ein AUG was ebenfalls innerhalb der kodierenden Sequenz des N-Terminus, aber außerhalb des Leserahmens lag, konnte dagegen nicht als Translationsstart fungieren, nachdem es durch Mutagenese in den Leserahmen gebracht worden war. Damit ist beim humanen GPR34 der Translationsstart mehrfach abgesichert, um eine Translation des Rezeptors zu garantieren. Die Ergebnisse zeigten aber auch, dass nicht nur das AUG-Motiv sondern auch die umgebende Sequenz festlegt, ob das AUG als Translationsstart genutzt werden kann. Ähnliche Befunde sind auch aus anderen Studien bekannt <sup>25</sup>. Die mehrfache Absicherung des Translationsstarts und damit einher gehende Maximierung der Translation unterstreicht die hohe biologische Relevanz des GPR34. Die dabei entstehenden verkürzten oder verlängerten N-Termini leisten zusätzlich einen Beitrag zur Diversität des Rezeptors auf Proteinebene.

### **5.3 Geringe allelische Variabilität des GPR34-Gens in der menschlichen Population**

Ebenfalls zur Rezeptordiversität beitragen können Allelvarianten eines Rezeptors. Die Variabilität in der kodierenden Nukleotidsequenz und eine nachfolgende mögliche Veränderung in der Aminosäuresequenz können strukturell und funktionell relevant sein. Der GPR34 ist im menschlichen Genom auf dem X-Chromosom kodiert. Damit können diese Veränderungen in der Aminosäuresequenz schnell populationsgenetisch in Erscheinung treten, da die männlichen Vertreter nur eine Genkopie besitzen. Von 105 weltweit untersuchten Individuen konnte ein männliches Individuum mit einem GTA- zu GTG-Austausch in der DNA Sequenz des GPR34 an Position 296 identifiziert werden. Diese Mutation führt jedoch zu keiner Veränderung in der Aminosäuresequenz (stille Mutation). Die ebenfalls untersuchte Mutter war in dieser Position heterozygot. Auch Iida und Nakamura <sup>26</sup> konnten in einer Untersuchung von 29 GPCR in 48 japanischen Individuen beim GPR34 keine SNP (single nucleotide polymorphisms) finden. Das belegt, dass die GPR34-Nukleotidsequenz innerhalb eines sehr stringent konservierten Bereiches liegt. Dies ist möglicherweise auch durch die Lage im Intron der Caseinkinase CASK begründet. Individuelle Variationen sind dabei nicht ausgeschlossen, wie der Fall des Europäers mit seinem GTA- zu GTG-Austausch an Position 296 belegt. Hochfrequente SNP kommen im humanen GPR34 nicht vor, was von einer geringen Alleldiversität zeugt. Deshalb ist davon auszugehen, dass Polymorphismen im

GPR34 (SNP mit einer Frequenz von >1 % in der Bevölkerung) nicht signifikant zur einer möglichen Funktionsdiversität des Rezeptors beitragen.

Die Schimpansen und Bonobo GPR34-kodierende Sequenz besitzt zwei stille Mutationen im Vergleich zur menschlichen Sequenz. Auch hier konnten keine Mutationen innerhalb der kodierenden GPR34-Sequenz zwischen Individuen nachgewiesen werden. Damit ist der GPR34 auch in den letzten 6,5 Mill. Jahren der menschlichen Entwicklung stark konserviert worden, was wiederum für eine essentielle Bedeutung des Rezeptors spricht.

#### **5.4 Ubiquitäre Expression von GPR34 und GPR34-like Rezeptoren**

Bereits kurz nach der Entdeckung des Rezeptors wurde von Marchese bzw. Schöneberg<sup>17,18</sup> in Maus und Mensch unabhängig voneinander eine ubiquitäre Expression des GPR34 in verschiedensten Geweben und Organen auf mRNA-Ebene durch „Northern Blot“ nachgewiesen. Dies konnte auch in den hier gezeigten Expressionsstudien in Maus, Ratte und Haushuhn nachvollzogen werden.

In den quantitativen Untersuchungen der Rezeptorexpression im Haushuhn (*Gallus gallus*) konnte erstmals ein Vergleich der Expression von GPR34 und GPR34-like gezeigt werden. Zum GPR34-like gab es bisher noch keine Expressionsdaten, da er in Säugern nicht oder nur als Pseudogen vorkommt. Im Haushuhn lässt sich feststellen, dass die GPR34-like-mRNA insgesamt niedriger exprimiert ist als die GPR34-mRNA und dass das Verhältnis der beiden Rezeptoren hauptsächlich durch Unterschiede in der GPR34-mRNA-Expression variabel ist.

Die bereits von Schöneberg et al. 1999<sup>18</sup> geäußerte Vermutung, dass der GPR34 ubiquitär exprimiert sein könnte, da er in vaskulärem bzw. lymphatischen Gewebe vorkommt, wurde durch Sugo et al. 2006 bestätigt<sup>21</sup>, indem sie eine sehr hohe spezifische Expression in Mastzellen nachwiesen, die bei ca. 15-40% der Expression von GAPDH lag. Es ist allerdings nicht möglich, Gewebe vollständig von Mastzellen zu befreien, um eine wirklich organspezifische Expression nachzuweisen. Deshalb wäre in zukünftigen Experimenten ein Mitführen einer Mastzell-spezifischen Kontrolle empfehlenswert, um den Anteil der Expression des GPR34 in Mastzellen am Ergebnis abschätzen zu können. Ob selbige Mastzellspezifität auch für den GPR34-like gilt, ist noch nicht nachgewiesen.

#### **5.5 Etabliertes Mausmodell ist vital und fortpflanzungsfähig**

Ein wichtiges Hilfsmittel zum Verständnis der Funktion eines „orphan“ GPCR ist die Generierung eines Rezeptor-defizienten Mausmodelles. Da im Laufe der Arbeiten zur

Promotion ein Agonist für den GPR34 durch Sugo et al. identifiziert werden konnte <sup>21</sup>, diese Ergebnisse aber auch wieder in Zweifel gezogen wurden <sup>27</sup>, ist dieses Mausmodell nun um so hilfreicher die Wirkung des Liganden auch *in vivo* zu testen.

Durch die Generierung eines GPR34-defizienten Mausmodelles konnte festgestellt werden, dass die Maus auch ohne GPR34 lebensfähig ist. Die GPR34-defizienten Mäuse zeigen unter Standardbedingungen ein normales Verhalten und in initialen Untersuchungen keine physiologischen Veränderungen. Eine genauere Charakterisierung des GPR34-knock-out-Mausmodells im Vergleich zu den Wildtyp-Mäusen wurde kürzlich publiziert <sup>27</sup>. Ein Fehlen des GPR34 führt in der Maus zu einer veränderten Immunantwort gegenüber Antigenen und Pathogenen. In dieser Arbeit konnte auch gezeigt werden, dass LysoPS nicht der Agonist am GPR34 der Maus ist.

## 6 Referenzen

1. Birnbaumer, L., Abramowitz, J. & Brown, A.M. Receptor-effector coupling by G proteins. *Biochim. Biophys. Acta* **1031**, 163-224 (1990).
2. Lefkowitz, R.J., Cotecchia, S., Samama, P. & Costa, T. Constitutive activity of receptors coupled to guanine nucleotide regulatory proteins. *Trends Pharmacol. Sci* **14**, 303-307 (1993).
3. Attwood, T.K. & Findlay, J.B. Fingerprinting G-protein-coupled receptors. *Protein Eng* **7**, 195-203 (1994).
4. Kolakowski, L.F. GCRDb: a G-protein-coupled receptor database. *Recept. Channels* **2**, 1-7 (1994).
5. Horn, F. et al. GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res* **26**, 275-279 (1998).
6. Bockaert, J. & Pin, J.P. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J* **18**, 1723-1729 (1999).
7. Fredriksson, R., Lagerström, M.C., Lundin, L.-G. & Schiöth, H.B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol* **63**, 1256-1272 (2003).
8. Davies, M.N. et al. Proteomic applications of automated GPCR classification. *Proteomics* **7**, 2800-2814 (2007).
9. Gloriam, D.E., Fredriksson, R. & Schiöth, H.B. The G protein-coupled receptor subset of the rat genome. *BMC Genomics* **8**, 338 (2007).
10. Vassilatis, D.K. et al. The G protein-coupled receptor repertoires of human and mouse. *Proc. Natl. Acad. Sci. U.S.A* **100**, 4903-4908 (2003).
11. Schöneberg, T. et al. Structural and functional evolution of the P2Y<sub>12</sub>-like receptor group. *Purinergic Signal* **3**, 255-268 (2007).
12. Abbracchio, M.P. et al. International Union of Pharmacology LVIII: update on the P2Y G protein-coupled nucleotide receptors: from molecular mechanisms and pathophysiology to therapy. *Pharmacol. Rev* **58**, 281-341 (2006).
13. Nonaka, Y., Hiramoto, T. & Fujita, N. Identification of endogenous surrogate ligands for human P2Y<sub>12</sub> receptors by in silico and in vitro methods. *Biochem. Biophys. Res. Commun* **337**, 281-288 (2005).
14. von Kügelgen, I. Pharmacological profiles of cloned mammalian P2Y-receptor subtypes. *Pharmacol. Ther* **110**, 415-432 (2006).
15. Paruchuri, S. et al. Leukotriene E4-induced pulmonary inflammation is mediated by the P2Y<sub>12</sub> receptor. *J. Exp. Med* **206**, 2543-2555 (2009).
16. Burnstock, G. Purinergic signalling--an overview. *Novartis Found. Symp* **276**, 26-48; discussion 48-57, 275-281 (2006).
17. Marchese, A. et al. Discovery of three novel orphan G-protein-coupled receptors. *Genomics* **56**, 12-21 (1999).
18. Schöneberg, T. et al. A novel subgroup of class I G-protein-coupled receptors. *Biochim. Biophys. Acta* **1446**, 57-70 (1999).
19. Schulz, A. & Schöneberg, T. The structural evolution of a P2Y-like G-protein-coupled receptor. *J. Biol. Chem* **278**, 35531-35541 (2003).
20. Kostenis, E., Waelbroeck, M. & Milligan, G. Techniques: promiscuous G $\alpha$  proteins in basic research and drug discovery. *Trends Pharmacol. Sci* **26**, 595-602 (2005).

21. Sugo, T. et al. Identification of a lysophosphatidylserine receptor on mast cells. *Biochem. Biophys. Res. Commun* **341**, 1078-1087 (2006).
22. Frasch, S.C. et al. NADPH oxidase-dependent generation of lysophosphatidylserine enhances clearance of activated and dying neutrophils via G2A. *J. Biol. Chem* **283**, 33736-33749 (2008).
23. Lloret, S. & Moreno, J.J.  $\text{Ca}^{2+}$  influx, phosphoinositide hydrolysis, and histamine release induced by lysophosphatidylserine in mast cells. *J. Cell. Physiol* **165**, 89-95 (1995).
24. Bruni, A. Autacoids from membrane phospholipids. *Pharmacol Res Commun* **20**, 529-544 (1988).
25. Kozak, M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* **299**, 1-34 (2002).
26. Iida, A. & Nakamura, Y. Identification of 156 novel SNPs in 29 genes encoding G-protein coupled receptors. *J. Hum. Genet* **50**, 182-191 (2005).
27. Liebscher, I. et al. Altered immune response in mice deficient for the G protein-coupled receptor GPR34. *J. Biol. Chem* **286**, 2101-2110 (2011).

## **Anlagen**

### **Erklärung über die eigenständige Abfassung der Arbeit**

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, und dass die vorgelegte Arbeit weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt an der Entstehung der vorliegenden Arbeit beteiligt waren.

.....

Datum

.....

Unterschrift

## Eigene Publikationen

1. Neuschäfer-Rube, F., **Engemaier, E.**, Koch, S., Böer, U. & Püschel, G.P. Identification by site-directed mutagenesis of amino acids contributing to ligand-binding specificity or signal transduction properties of the human FP prostanoid receptor. *Biochem. J* **371**, 443-449 (2003).
2. **Engemaier, E.**, Römpler, H., Schöneberg, T. & Schulz, A. Genomic and supragenomic structure of the nucleotide-like G-protein-coupled receptor GPR34. *Genomics* **87**, 254-264 (2006).
3. Schöneberg, T., Hermsdorf, T., **Engemaier, E.**, Engel, K., Liebscher, I., Thor, D., Zierau, K., Römpler, H. & Schulz, A. Structural and functional evolution of the P2Y(12)-like receptor group. *Purinergic Signal* **3**, 255-268 (2007).
4. Liebscher, I., Müller, U., Teupser, D., **Engemaier, E.**, Engel, K.M., Ritscher, L., Thor, D., Sangkuhl, K., Ricken, A., Wurm, A., Piehler, D., Schmutzler, S., Fuhrmann, H., Albert, F.W., Reichenbach, A., Thiery, J., Schöneberg, T. & Schulz, A. Altered immune response in mice deficient for the G protein-coupled receptor GPR34. *J. Biol. Chem* **286**, 2101-2110 (2011)
5. Ritscher, L., **Engemaier, E.**, Stäubert C., Liebscher I., Römpler H., Schulz A.& Schöneberg T. Determinants involved in agonist specificity of the G protein-coupled receptor GPR34. *J. Biol. Chem* (in revision)

## Danksagung

In großer Dankbarkeit bin ich Prof. Dr. Torsten Schöneberg als meinem Doktorvater verbunden, der mir immer geduldig, inspirierend und unermüdlich zur Seite stand und steht.

Sehr dankbar bin ich auch Dr. Holger Römpler, der viel Kraft und Zeit in mich und meine Projekte investiert hat, die Welt menschlich und verständnisvoll sieht und sich dabei manchmal selbst etwas zuviel abverlangt. Danken möchte ich ebenfalls Dr. Angela Schulz für die gute Betreuung im Labor - mit Ihrer Hilfe konnte ich mein Methodenspektrum entscheidend erweitern.

Ein ganz herzlicher Dank gilt dem gesamten Team der Molekularen Biochemie in dem ich sehr liebe, wertvolle Menschen kennenlernen durfte und die diese Arbeit erst möglich gemacht haben: Frau Kerstin Rückauer, Dr. Thomas Hermsdorf, Dr. Katrin Sangkuhl, Dr. Doreen Thor, Dr. Anke Edelmann, Dr. Katrin Tanneberger, Dr. Verena Pfeiffer, Dr. Wibke Busch, Dr. Ines Liebscher, Dr. Claudia Stäubert, Dr. Rainer Strotmann, Iris Bösel, Stine Beyer und Petra Krumbholz – vielen Dank!

Mein weiterer Dank gilt den Ehemaligen des Graduiertenkollegs „Vaskuläre Regulationsmechanismen“ der FU Berlin, die mir immer mit Rat und Tat zur Seite standen und unter denen ich sehr gute Freunde gefunden habe.

Last but not least – ein ganz dickes Dankeschön an meine Freunde und Familie, die unentwegt für mich da sind und die die Basis für diese Arbeit gelegt haben und legen.